One Is All, All Is One:
Cross-Modal Displays for Inclusive Design and Technology

Esenkaya, Tayfun

Award date:
2020

Awarding institution:
University of Bath

One Is All, All Is One:

Cross-Modal Displays for Inclusive Design and Technology

Tayfun Esenkaya

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Horizon 2020: Marie Skłodowska-Curie Fellowship with Industrial Research Enhancement

Centre for Digital Entertainment

Cross-Modal Cognition Lab

Department of Computer Science

November 2019

COPYRIGHT

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of Experiment I of Chapter II where 100% of data collection and development of Haptic Face Display were carried out by other researchers.

To anaduk, babaduk, arı, kelebek, lili & loushi.
Without their deepest support, this thesis wouldn't come together.

Contents

COPYRIGHT 3

Declaration of any previous submission of the work 3

Declaration of authorship 3

Contents 5

List of Figures, Graphs, Boxplots 10

List of Tables 12

ACKNOWLEDGEMENTS 13

Abstract 14

Reflections I 15

CHAPTER I

Potential Inclusive Applications of Sensory Substitution Techniques for all of 'Us' 18

Declaration 19

Acknowledgements 20

1.0 Abstract 21

1.1 Introduction 22

1.1.1 From Them to Us: Change of Perspective Taking 22

1.1.2 Us vs. Them: Contemporary Perspectives from Technology Industry 23

1.1.3 Socioeconomic Impact of Inclusion 24

1.1.3.1 A Historical Example 24

1.1.3.2 The Business Case 25

1.1.4 The Inclusive Design Mindset 26

1.1.5 From Predictive Text to SMS to Sensory Substitution Techniques 27

1.2 An Overview of Sensory Substitution Phenomena 28

1.2.1 Sensory Substitution 28

1.2.2 Sensory Substitution Devices as Cross-Modal Displays 29

1.2.2.1 Empirical Evidence from Applied Auditory-to-Visual and Tactile-to-Visual Sensory Substitution Techniques 30

1.2.2.2 Widespread Adoption Problems of Sensory Substitution Techniques 32

1.2.3 Potential Applications of Cross-Modal Displays for all of Us 33

1.3 Conclusion 34

1.4 References 36

Reflections II 45

CHAPTER II

Cross-Modal Tactile and Sonification Associations to Enrich Digital Emotion Communication 47

Declaration 48

Acknowledgements 49

2.0 Abstract 50

2.1 Introduction 51

2.2 Experimental Investigation 54

2.3 Methods 57

2.3.1 Participants 57

2.3.2 Apparatus 57
2.3.2.1 Haptic Face Display (HFD) 57
2.3.2.2 The vOICe 59

2.3.3 Stimuli 59
2.3.3.1 Stimulation Patterns 59

2.3.4 Presentation of the Feedback 62
2.3.4.1 Experiment I: Tested Associations between Cross-Modal Tactile Feedback and Basic Emotions 62
2.3.4.2 Experiment II: Tested Associations between Cross-Modal Sonification Feedback and Basic Emotions 62
2.3.4.3 Experiment III: Tested Associations between Visual Feedback and Basic Emotions 62

2.3.5 Experiment 63
2.3.5.1 Experimental Procedure 63

2.4 Results 64

2.4.1 Strong Associations between Emotional Responses and Stimulation Patterns 64

2.4.2 Emotional Intensity 65
2.4.2.1 Within Tactile Feedback 65
2.4.2.2 Within Sonification Feedback 66
2.4.2.3 Within Visual Feedback 68
2.4.2.4 Overall Comparisons 68

2.5 Discussion 69

2.5.1 Spatial Cross-modal Associations with Basic Emotions 71

2.5.2 Establishing a Framework of Cross-Modal Associations, Limitations and Future Perspectives 73

2.6 Conclusion 76

2.7 References 77

2.8 Appendices 81

Appendix 2.A: Polar Maps of Stimulation Patterns with respect to Six Spatial Factors *81*

Appendix 2.B: Contingency Table of Response Frequencies *84*

Reflections III 85

CHAPTER III

Two Is Better Than One: Multisensory Combination of Auditory and Tactile Cross-Modal Displays 87

Declaration 88

Acknowledgements 89

3.0 Abstract 90

3.1 Introduction 91

3.2 Background 92

3.2.1 A Framework for Multisensory Processing *93*

3.2.1.1 Principles of Spatial and Temporal Coincidence 94

3.2.1.2 Principle of Inverse Effectiveness 94

3.2.1.3 Multisensory Integration 95

3.2.1.4 Multisensory Combination 96

3.2.2 Sensory Substitution *97*

3.2.2.1 Sensory Substitution Devices as Cross-Modal Displays 98

3.2.2.2 Multisensory Use of Auditory and Tactile Cross-Modal Displays 100

3.3 Experimental Investigation 102

3.3.1 Experimental Hypotheses *104*

3.4 Methods 104

3.4.1 Participants *104*

3.4.2 Cross-Modal Box *104*

3.4.3 Stimuli *105*

3.4.4 Experimental Conditions *108*

3.4.5 Procedure *108*

3.4.5.1 Training 108

3.4.5.2 Experiment 109

3.5 Results 109

3.5.1 Calculating Response Accuracy and Reaction Times *110*

3.5.1.1 Analysing Response Accuracy 110

3.5.1.2 Reaction Times 111

3.5.2 Gender Differences *112*

3.5.2.1 Response Accuracy 112

3.5.2.2 Reaction Times 112

3.5.3 Task and Display Mode Difficulty *113*

3.5.3.1 Task Difficulty 113

3.5.3.2 Display Mode Difficulty 113

3.5.4 Qualitative Strategies *114*

3.5.4.1 Auditory Mode ... 114
3.5.4.2 Tactile Mode ... 114
3.5.4.3 Multisensory Mode ... 115

3.6 Discussion ... 115

3.6.1 Summary of Results ... 115

3.6.2 Theoretical Implications ... 117
3.6.2.1 Limitations and Future Perspectives ... 118

3.6.3 Practical Implications ... 120

3.7 Conclusion ... 121

3.8 References ... 123

3.9 Appendix ... 129

Appendix 3.A: Training Stimuli ... 129

Appendix 3.B: Experimental Stimuli ... 129

Appendix 3.C: Cross Tabulation of Response Frequencies and Stimuli ... 130
3.C.1 Letter Recognition with Auditory Mode ... 130
3.C.2 Shape Recognition with Auditory Mode ... 130
3.C.3 Letter Recognition with Tactile Mode ... 131
3.C.4 Shape Recognition with Tactile Mode ... 131
3.C.5 Letter Recognition with Multisensory Mode ... 132
3.C.6 Shape Recognition with Multisensory Mode ... 132

Reflections IV ... 133

CHAPTER IV

Auditory and Tactile Cross-Modal Displays Can Help with the Last 10 Metres of Navigation ... 135

Declaration ... 136

Acknowledgements ... 137

4.0 Abstract ... 138

4.1 Introduction ... 139

4.2 Background ... 142

4.2.1 Cross-Modal Displays ... 142

4.2.2 Multisensory Combination ... 143

4.3 Experimental Investigation ... 144

4.3.1 Experimental Hypotheses ... 146

4.4 Methods ... 146

4.4.1 Participants ... 146

4.4.2 Apparatus ... 147
4.4.2.1 Cross-Modal Display Prototypes ... 147
4.4.2.2 Motion Tracking ... 150

4.4.3 Stimulus Design ... *150*
4.4.3.1 Bird's-eye View Maps ... 151
4.4.4 Triangle Completion Task ... *151*
4.4.5 Experimental Conditions ... *152*
4.4.5.1 Experiment I: Tested Kinaesthesia, Sonification and Sonification-Kinaesthesia Display Modes from First-Person Perspective ... 154
4.4.5.2 Experiment II: Tested Tactile, Sonification and Sonification-Tactile Display Modes from Bird's-Eye View Perspective ... 154
4.4.6 Experimental Procedure ... *154*
4.4.6.1 Training ... 155
4.4.6.2 Navigation Experiment ... 155
4.5 Results ... 156
4.5.1 Experiment I: Tested Kinaesthesia, Sonification and Sonification-Kinaesthesia Display Modes from First-Person Perspective ... *157*
4.5.1.1 Constant Error ... 157
4.5.1.2 Variable Error ... 158
4.5.1.3 The Effect of Learning ... 159
4.5.1.4 Qualitative Strategies ... 160
4.5.2 Experiment II: Tested Tactile, Sonification and Sonification-Tactile Display Modes from Bird's-Eye View Perspective ... *161*
4.5.2.1 Constant Error ... 161
4.5.2.2 Variable Error ... 162
4.5.2.3 The Effect of Learning ... 163
4.5.2.4 Qualitative Strategies ... 164
4.5.3 Between-Subject Comparison between First-Person and Bird's-eye View Perspectives ... *165*
4.6 Discussion ... 165
4.6.1 Performance of Sonification-Kinaesthesia Prototype ... *166*
4.6.2 Performance of Sonification-Tactile Prototype ... *168*
4.6.3 Performance of Viewing Perspectives ... *170*
4.6.4 Future Perspectives ... *171*
4.6.4.1 Theoretical Implications and Limitations ... 171
4.6.4.2 Practical Implications and Limitations ... 173
4.7 Conclusion ... 175
4.8 References ... 177
Reflections V ... 183
Closing Summary ... 186
References ... *189*

List of Figures, Graphs, Boxplots

Figure 2.1a (left) displays the Haptic Face Display mounted on the back of an ergonomic mesh chair. **Figure 2.1b** (right) show a single tactor strip embedded on the HFD. 58

Figure 2.2 represents the stimulation patterns frame by frame. They are named arbitrarily for convenience. Each frame represents part of a linear sequence of the stimulation patterns from left to right and is presented for 500ms via each display mode. 61

Figure 2.3a (left) shows the contingency table of strong associations between emotional responses and stimulation patterns with respect to each display mode (tactile, sonification, visual). These frequencies are provided in percentages of total responses given to a stimulation pattern associated with an emotional response. Lightly shaded cells indicate strong associations. Additionally, percentages of overall emotional response frequencies are presented next to the emotional response. **Figure 2.3b** (right) displays the biplots representing associations between stimulation patterns (green) and emotion types (purple) with respect to each display mode. Some stimulation pattern names are shortened (e.g. 'Alternate Top-Bottom' is 'alt tb' for aesthetics reasons). 67

Figure 2.4 displays the polar maps of emotional responses, which were created by overlapping the polar map of each stimulation pattern that was strongly associated with the given emotion. Green, orange and grey respectively represents sonification, tactile and visual display modes. Darker connections indicate areas where there are more overlaps. 72

Figure 3.1 shows a user with Cross-Modal Box in the multisensory mode. The user has the intra-oral interface on her tongue for electrotactile cues and bone conduction headphones for sonifications. 106

Figure 3.2 exemplifies a shape (i.e. circle) and a letter (i.e K) stimuli. Images on the right (b & d) display the tactile representations conveyed via the intra-oral interface. These images were saved and then sonified with The vOICe to create the auditory stimuli. The spectrograms on the left (a & c) represent the spectrum of frequencies of these sonifications. 107

Graph 3.1 represents accuracy with respect to the display mode and stimuli type. Error bars show +/- 1 SE. The average correct scores for letter stimuli via auditory, tactile and multisensory modes were respectively 40.4% (SE = 4.0), 67.5% (SE = 3.1) and 78.1% (SE = 3.9). The average scores for shape stimuli were respectively 47.7% (SE = 3.5), 61.2% (SE = 3.7), 76.8% (SE = 3.2). 111

Graph 3.2 shows the mean reaction times in seconds in relation to the display mode and stimuli type. Error bars show +/- 1 SE. The average reaction times for letter stimuli via auditory, tactile and multisensory modes were respectively 14.4s (SE = 1.0), 16.6s (SE = 1.5) and 14.5s (SE = 1.0). The average reaction times for shape stimuli were respectively 13.3s (SE = 0.9), 14.5s (SE = 0.9) and 13.6s (SE = 0.9). 112

Figure 4.1 displays a user wearing Sonification-Kinaesthesia Prototype. 148

Figure 4.2 displays a user using the Sonification-Tactile Prototype in the multisensory mode. 150

Figure 4.3 (left) and **Figure 4.4** (right) display the three targets and their triangular configuration used in the experiments. The start point indicates the point where participants started exploring the targets (exploration phase) and also started the navigation tasks (task phase). Blue-taped areas on the floor show the training configuration of the targets while the start point is invisible. 151

Figure 4.5 illustrates the birds-view maps of training configuration (left) and experimental configuration (right). The start points are respectively on the bottom left and the right corners. 151

Figure 4.6 illustrates the egocentric route (left) and allocentric route (right) of the triangle completion task. 152

Boxplot 4.1 shows constant errors in metres for kinaesthesia, sonification and sonification-kinaesthesia display modes with respect to egocentric (cyan) and allocentric (magenta) routes. 157

Boxplot 4.2 shows variable errors in metres for kinaesthesia, sonification and sonification-kinaesthesia display modes with respect to egocentric (cyan) and allocentric (magenta) routes. 158

Graph 4.1 represents the effect of learning across 10 trials for egocentric route (left) and allocentric route (right). The display modes are indicated by black (sonification-kinaesthesia), magenta (sonification), and cyan (kinaesthesia). 159

Boxplot 4.3 shows constant errors in metres for tactile, sonification and sonification-tactile display modes with respect to egocentric (cyan) and allocentric (magenta) routes..... 161

Boxplot 4.4 shows variable errors in metres for tactile, sonification and sonification-tactile display modes with respect to egocentric (cyan) and allocentric (magenta) routes..... 162

Graph 4.2 represents the effect of learning across 10 trials for egocentric route (left) and allocentric route (right). The display modes are indicated by black (sonification-tactile), magenta (sonification) and cyan (tactile). 163

List of Tables

Table 2.1 summarises the average (∓ SD) of emotional intensities with respect to display mode and emotion type. 68

Table 3.1 summarises the percentages of responses given to the task difficulty questions, where the elements in the first row were asked to be ordered from the easiest to the hardest for each element in the first column. Cells are shaded with respect to the value of percentages (i.e. 100% is pitch black and 0% is white). ..... **Error! Bookmark not defined.**

Table 3.2 summarises the percentages of responses given to the display mode difficulty questions, where participants were asked to order the display modes from the easiest to use to the hardest. Cells are shaded with respect to the value of percentages (i.e. 100% is pitch black and 0% is white)..... 113

Table 4.1 Table to show the procedures used to present each display mode in all phases of the first and second experiment..... 153

ACKNOWLEDGEMENTS

This research has been funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665992, and the UK's EPSRC Centre for Doctoral Training in Digital Entertainment (CDE), EP/L016540/1.

Abstract

Sensory substitution phenomena transform the representation of one sensory form into an equivalent from a different sensory origin. For example, a visual feed from a camera can be turned into something that can be touched or sounds that can be heard. The immediate applications of this can be seen in developing assistive technologies that aid vestibular problems, and visual and hearing impairments. This raises the question of whether perception with sensory substitution is processed like an image, or like a surface, or a sound. Sensory substitution techniques offer a great opportunity to dissociate the stimulus, the task and sensory modality, and thus provide a novel way to explore the level of representation that is most crucial for cognition. Accordingly, state-of-the-art sensory substitution techniques contribute significantly to the understanding of how the brain processes sensory information and also represents it with distinct qualia. This progressively advances cognitive theories with respect to multisensory perception and cross-modal cognition. Due to its versatility, sensory substitution phenomena also carry the applications of cognitive theories to other interdisciplinary research areas such as human-computer interactions (HCI). In HCI, cross-modal displays utilise sensory substitution techniques to augment users by enabling them to acquire sensory information via a sensory channel of different origin. The modular and flexible nature of cross-modal displays provide a supplementary framework that can appeal to a wider range of people whose physical and cognitive capabilities vary on a continuum. The present thesis focuses on the inclusive applications of sensory substitution techniques and cross-modal displays. Chapter I outlines the inclusive design mindset and proposes a case for applications of sensory substitution techniques for all of us. Chapter II and Chapter IV evaluates cross-modal displays in digital emotion communication and navigation applications respectively. Chapter III offers a methodology to study sensory substitution in a multisensory context. The present thesis evidences that perception with cross-modal displays utilises the capabilities of various senses. It further investigates the implication of this and suggests that cross-modal displays can benefit from multisensory combination. With multisensory combination, cross-modal displays with unisensory and multisensory modes can deliver complementary feedback. In this way, it is argued users can gain access to the same inclusive information technology with customised sensory channels. Overall, the scope of the present thesis approaches sensory substitution phenomena from an HCI perspective with theoretical implications grounded in cognitive sciences.

Reflections I

"We are still information hunter-gatherers."
Eleanor Gibson

I would like to thank those who will read this thesis and hope that you will find an interesting insight or two. Unlike the traditional format, the current thesis is submitted in an alternative composition. It means that each chapter constitutes a manuscript ready for submission, or already submitted, to a peer-reviewed journal in human-computer interaction (HCI) research. Each chapter is therefore stand-alone, and hence might share some of the background content. Between the chapters, where appropriate, there are also '*Reflection*' sections. *Reflections* are intended to summarise what has been discussed previously and to inform the reader what is coming next.

Reflections I briefly introduces my research interests and how this thesis came together over the last 4 years. In the Horizon 2020 Marie Skłodowska-Curie Fellowship for Industrial Research Enhancement (FIRE) programme, I gained invaluable opportunities to train myself in a breadth of multidisciplinary fields. Having previously studied molecular biology, genetics and bioengineering at undergraduate level, how the genetic information led to the human experience and its qualia fascinated me. The FIRE programme offered me the intellectual freedom to inquire how information is born and gains meaning. There was a significant taught component in Year 1, which included an array of master-level courses in entrepreneurship and innovation, business and management, industry, research and development, and public

engagement. From Year 2 onwards, I have started specialising in topics of research interest in HCI. How we interact with technology has been just as intriguing to me as how we interact with sensory information from a cognitive/neuroscience perspective. One of the research areas where these two perspectives intersect is sensory substitution. From an HCI perspective, sensory substitution techniques could be applicable in display mode developments by enabling interaction with a novel form of information via an intact sensory organ. Understanding how sensory substitution works could also contribute to the cognitive and neural theories of the brain and how we interact with sensory information.

The current applications of sensory substitution techniques concern the development of assistive technologies, especially for those with visual impairments. There is a vast amount of research that evidences how sensory substitution can help the acquisition of visual information via the sensations of hearing and touch. Others also demonstrate that sensory substitution techniques can be applied in the rehabilitation of hearing loss and vestibular impairments. Nevertheless, sensory substitution devices as assistive technologies are not yet widely adopted. One reason for this may be because sensory substitution is often taken for its literal meaning (i.e. a phenomenological substitution between sensory forms). It is therefore lost in translation between interdisciplinary research.

Another line of research, relatively fewer in numbers, examines sensory substitution techniques in developing novel forms of cross-modal display modes that can enrich our interactions with the digital world. In the inclusion context, this area of research benefits from an approach to sensory substitution that utilises the brain's ability to portray formless neural signals with meaningful representations. With this approach, sensory substitution techniques can be deployed to develop new types of interaction with technology, independent of the sensory form of their interfaces. This would be possible as cross-modal display modes can be substituted in a way meaningful to the user via different sensory channels without requiring specialist hardware. In this way, the same technology can be made available to a wider range of people. This requires change in thinking.

Accordingly, the first chapter of the present thesis reviews how inclusive design mindset is different than the traditional usability and accessibility frameworks with contemporary examples of social, industrial, policy, and research and development cases. It then outlines a case for the potential inclusive applications of sensory substitution techniques. By doing so, it aims to promote sensory substitution phenomena to a wider circle of researchers to develop inclusive technologies in a variety of use cases. Chapter II and Chapter IV demonstrate various inclusive use cases of sensory substitution techniques in digital emotion communication and navigation applications respectively. Chapter III additionally offers a methodology to study sensory substitution techniques in a multisensory context and suggests possible inclusive applications in extended reality platforms and tangible interactions. Overall, the scope of the present thesis covers sensory substitution phenomena from an HCI perspective with theoretical implications grounded in cognitive sciences. With an inclusive design mindset, cross-modal display modes that empower sensory substitution techniques can be improved to benefit mainstream applications, while simultaneously expanding their widespread adoption as an assistive technology.

CHAPTER I

Potential Inclusive Applications of Sensory Substitution Techniques for all of ‘Us’

Declaration

<table>
<tr><td colspan="6">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="6">Potential Inclusive Applications of Sensory Substitution Techniques for all of ‘Us’</td></tr>
<tr><td colspan="6">Publication status (tick one)
Draft manuscript [x] Submitted [ ] In review [ ] Accepted [ ] Published [ ]</td></tr>
<tr><td>Publication details (reference)</td><td colspan="5">N/A</td></tr>
<tr><td colspan="6">Copyright status (tick the appropriate statement)
I hold the copyright for this material [x] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]</td></tr>
<tr><td>Candidate’s contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="5">The candidate predominantly executed the
Formulation of ideas: %90
Design of methodology: N/A
Experimental work: N/A
Presentation of data in journal format: %90
For details, please see acknowledgements on the next page</td></tr>
<tr><td>Statement from Candidate</td><td colspan="5">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td colspan="2"></td><td>Date</td><td colspan="2"></td></tr>
</table>

Acknowledgements

I am grateful to Vanessa Lloyd-Esenkaya and Bora Esenkaya for their insightful conversations. I also thank Michael Proulx for his advice and resourceful supervision throughout the process.

1.0 Abstract

Sensory substitution techniques are a perceptual and cognitive phenomena, and are used to represent one sensory form with a novel alternative. The status quo of sensory substitution techniques is currently focused on the development of assistive technologies whereby visually impaired users could acquire visual information via auditory and tactile cross-modal feedback. Despite their evidenced success in scientific research, sensory substitution techniques have not yet gained widespread adoption amongst the visually impaired populations. The current review will therefore explore why this might be the case from an inclusive design perspective with contemporary examples of socioeconomic impact, and argue that shifting the focus from assistive to mainstream applications might resolve some of these issues.

1.1 Introduction

Inclusive design, also known as design for all (DfA) in Europe and universal design in USA and Japan, is about the acceptance of and tolerance for difference by involving everyone and ensuring that everyone has a sense of belonging (for detailed reviews, see [34,42,47]). It is distinguished in essence from the integration design rhetoric. Integration deploys accessibility concepts, for example within social organisations [144], built environments [66] or technology products [150], to those with special physical and/or cognitive impairments [144]. Integration therefore assumes that the delivery of accessibility will enable those at disadvantage to fit into mainstream settings. In this way, integration is tied to the fundamental expectation that individual groups should prove their readiness in a mainstream setting [144]. Inclusion, on the other hand, has the inverse expectation: the mainstream setting should prove its readiness to accept every person [88,144]. The current review will give an overview of the way inclusive design can be applied to contemporary technologies which serve the broader population whose ability varies on a complex and continuous spectrum, rather than on singular and discrete categories [58,67,150,151]. The first part of this review will explain how traditional legacy frameworks, of usability and accessibility, can make the transition to more inclusive designs. Contemporary examples of the socioeconomic impact of this transition process will be provided, in the context of policies, business cases, and research and development resources. The second part of this review will outline sensory substitution techniques and argue a case for their implementation in the development of inclusive technologies, such as cross-modal displays.

1.1.1 From Them to Us: Change of Perspective Taking

With developments in national and international policies, such as the Americans with Disabilities Act 1990 [2], the Disability Discrimination Act 1995 [39], and the Equality Act [147], the focus in design decisions has gradually oriented from ‘THEM’ to ‘US’ [34]. This change in view has had a tremendous positive influence on how the physical and cognitive limitations/impairments of people were once conceived as congenital

or acquired incapacities/disabilities. This conception has progressively been metamorphosising into a social model in which indifferent design, services, environments and stereotypes are blamed for the segregation between able-bodiedness and unable-bodiedness [34]. In other words, it is the design decisions that are enabling and disabling, or including and excluding, our physical and cognitive capacities. Accessibility and usability should therefore not be viewed as separate entities. Rather, they should jointly form the intrinsic qualities of emerging tools to develop, market and commercialise inclusive technologies, art forms, institutions and such. This changing perspective by no means denies the incredible impact that assistive technologies can have on those who need them. These premises learn from the past rather than criticising it and look forward to the future with great optimism. With inclusive design, "if everyone is not special, maybe you can be what you want to be" [122].

1.1.2 Us vs. Them: Contemporary Perspectives from Technology Industry

Within the technology industry, Microsoft explicitly brought forward many unforeseen issues with traditional accessibility and usability frameworks by identifying the social and commercial potential that inclusive approaches to design can have. Microsoft reports, "… the concept of 'disability' may have limited the understanding of the need for accessible technology. Instead of assuming that accessible technology is only useful to a distinct group of people with disabilities, the IT industry must consider the wide range of people who could benefit…" [138]. In recent years, this issue has been polarised to the extent that product developers and companies have perceived mainstream products to be for the able-bodied only, and the only way to serve the unable-bodied population is to create specialist solutions (i.e. assistive technologies) that are designed with a disability-centric approach [29]. These arbitrary singular categorisations force design decisions to be exclusive and disabling, which inevitably results in many specialist solutions.

As a consequence, mainstream commercial organisations lose interest in the minority market, which unavoidably limits the supplies for research and development into

technologies that have the potential to empower a wide range of people [34,150]. With the division of available resources between mainstream and assistive technologies, the latter cannot be maintained and is therefore abandoned by those who use them [112]. Such division in resources also has a detrimental impact on mainstream technologies because people vary in complex ways on a wide continuous spectrum [58,67,151]. Indeed, research indicates that as many as two thirds of the population experience difficulties and frustration with mainstream technological tools [29]. Design decisions that segregate accessibility from inclusion therefore silently diminish the overall socioeconomic reach and impact that business, academic and social cases might otherwise have. This highlights the importance of inclusive design.

1.1.3 Socioeconomic Impact of Inclusion

1.1.3.1 A Historical Example

In the context of technology, a successful demonstration of how transitioning from legacy frameworks of usability and accessibility to inclusion can benefit different parties comes from the mainstream adoption of predictive text recognition. Predictive text simply enables one button to represent multiple keys (e.g. alphabetical input via the numeric keypad of mobile phones) with each key press resulting in a prediction of the next. It is now a widely used feature of mobile phones, but it was first functionalised to substitute assistive telecommunication devices for the deaf (TDD) [53]. TDD, also commonly known as teletypewriter, textphone or minicom, is an electronic device designed particularly for deaf people [99], enabling them to communicate over the telephone without the use of speech. Despite their benefits, TDDs were only available to a minority of the deaf population at the time of their inception, due to high costs [154]. This reduced the potential for deaf people to connect with others [109]. Once installed with predictive text recognition capabilities, however, standard telephones were equipped with a functionality similar to TDD [53]. This new-generation assistive technology consequently enabled deaf people to take independence and communicate freely with others without a significant financial cost.

Subsequently, mobile phone producers commercialised the use cases of predictive text recognition across an array of information and communication technologies, such as short message service (SMS), demanded by the mainstream population. In return, once dependent on TDDs to communicate with others, deaf people started using SMSs too, which shortly became popular among their communities [114]. 90% of deaf respondents of a survey, which measured their telecommunication preferences, agreed that the reason SMS proved so popular among them was because it could be used anywhere, without specialist equipment [113]. The commercial journey predictive text recognition has taken, from an assistive technology to a mainstream application, demonstrates how the development of inclusive technologies can have a positive socioeconomic impact, while simultaneously improving experiences for minority populations who were originally served by the assistive technology.

1.1.3.2 The Business Case

It is projected that the traditional integration rhetoric can scarcely maintain the status quo of mainstream settings, even if their economic models are sustainable [144]. This would undeniably result in an increasingly expensive specialised alternative sector for those with impairments as inclusion would be underfunded [144]. With inclusion, however, this scenario is more optimistic in the sense that it rearranges resources to proactively support acceptance of and tolerance for everyone [144,145]. While these proposals were initially discussed in the context of education, it is easy to identify the parallels it could share with the technology industry, when considering the case of commercialised technologies like predictive text recognition. This business case evidently outlines how resources for research and development can be purposely allocated to optimise the development cycle of innovative information and communication technologies.

It is rather a myth that minority user profiles have little economic significance [34]. On the contrary, a business case should be intrinsically inclusive for the very reason that it means capturing a greater market share [67] along with a definite competitive edge [150], increased net profit [78] and reduced service costs [96]. Greater market size

scales the potential for research and development of a product that can reach to a wider range of people. This consequently defines how much a technology could be adopted and commercialised, hence the greater socioeconomic impact. Inclusivity benefits everyone.

1.1.4 The Inclusive Design Mindset

Seeing through the myth, in which minority user profiles offer little economic significance and the unable-bodied cannot use mainstream technologies, calls for innovation that challenges assumptions about accessibility and usability in design thinking and decisions. In this respect, an inclusive design mindset matters significantly for a greater socioeconomical impact. Nevertheless, inclusive design neither commands a singular solution for all nor ignores the specialist solutions required in certain use cases [150]. It rather suggests that design should be flexible and modular for the best coverage of diversity and universal access [150].

In HCI research, this is achieved with two complementary scopes: the inclusive design mindset and the design process itself [110]. For this, participatory design is a well-suited design process that supports the full cooperation between the users and the development team during the end-to-end life cycle of the product development [124]. For example, the design process could begin with a focus group with the end users who are then iteratively consulted at every stage of the development process. When applied with the inclusive design mindset, participatory design welcomes a wider range of users in the process of designing, developing and commercialising technology products. In this respect, participatory design processes and the inclusive design mindset further advocate that accessibility is less about physical and cognitive disabilities and more about enabling universal access to technology [110].

Intuitively, this can be achieved by identifying the mutual physical and cognitive capacities of greater diversity by including a mixed group of users early in the participatory design process. Once recognised, these can be utilised to empower inclusive mainstream products. For example, the widespread adoption of SMSs by

mainstream and deaf populations was simply possible because both groups were able to see the phone display, type a text and comprehend language. Additionally, it was noted that the deaf community was able to adopt to SMS successfully as it did not require a special equipment for transmission unlike other unaffordable assistive technologies. The demand to communicate with others portrayed a significant business opportunity with a diverse yet larger market. As the result of unified research and development, affordable mobile phones with SMS capability was made available to a wider group of people. In contrast, SMS use, for example, would be still challenging and perhaps exclusive for the visually impaired population until screen readers were developed and embedded in information and communication technologies. However, it could be argued that the widespread adoption of phones in the first place made telecommunication more affordable and easier than the alternative written forms of correspondences, such as letters, for both mainstream and blind populations [19]. A similar technology behind screen readers could be inclusively utilised for audio books that would eventually benefit a greater population, blind and sighted alike. Another relevant example could be the invention and global adoption of the Braille writing system among the visually impaired populations, which was originally developed as a military application (i.e. its predecessor night writing) [111]. Braille was the first writing system with binary encoding, and its 6-bit code and digital nature resemble the modern electronic circuits that are used in digital devices [37,111]. In short, inclusive innovations are cyclic and reciprocal, benefitting a diverse group of people at each iteration. With the application of an inclusive participatory design process, these benefits to the society can be more interlinked.

1.1.5 From Predictive Text to SMS to Sensory Substitution Techniques

It is through the inclusion of a wider range of capabilities from a wider population that enables inclusive design to flourish and this, in turn, establishes a more inclusive society. The main message of inclusive design is to innovate beyond the traditional models of usability and accessibility criteria that unintentionally divides the research and development resources in commercial [18], academic and social organisations (e.g. educational [63] and public institutions [102]). In the remainder of this review,

we would like to argue a similar case (e.g. that of predictive text) for sensory substitution techniques, which are primarily utilised in assistive technologies targeted at visually impaired populations, and why they can be used for building inclusive technologies. In this way, we aim to unify and promote the research and development resources dedicated for sensory substitution. We will revisit its potential applications as cross-modal displays for all of us with an inclusive mindset.

1.2 An Overview of Sensory Substitution Phenomena

1.2.1 Sensory Substitution

Sensory substitution is a perceptual and cognitive phenomena in which some features of a sensory experience, such as of seeing, can be represented with a sensory form different than its sensory origin [50]. This is possible on two assertions: (i) 'concepts' (e.g. physical objects) of the environment share mutual features (e.g. spatial properties) when represented via different sensory origins, and (ii) our brains have evolved to perceive and process them in a coherent multisensory fashion [56,86,97,131,133,137]. Traditionally, it was conceived that the brain consisted of independent unisensory modules, which operated extensively (e.g. bottom-up facilitations) before multisensory percepts occurred [33,56]. This view was challenged whereby the second assertion has been further evidenced to execute metamodal computations and tasks (cognitive forms) as an integrated network, namely the metamodal organisation of the brain [107]. Intuitively, perceiving the edge of a cube, for example, might involve a mutual cognitive form via visual and/or tactile sensations irrespective of their sensory forms. In fact, the metamodal hypothesis has been repeatedly supported by a growing body of empirical evidence [20,56,104,119,120]. Furthermore, the bottom-up sensory responses are shown to be modulated by top-down facilitations (e.g. memory and attention) such that previously acquired associations and learning can enhance task-relevant multisensory responses (for a detailed review, see [33]).

Considering the metamodal organisation of the brain, sensory information is learnt and hence gradually associated with one another in relation to bottom-up and top-down facilitations. This is evidenced to the extent that boulders are conceived to be sour, lemons to be fast, prunes to be slow and the colour red to be heavy [157], with the latter example further existing even among the early blind who never experienced colour perception [12]. These abstract associations are argued to be the result of mutual conceptual dimensions between sensory information, and therefore shared across cultures and languages [12,131,134,157]. This immediately raises the question of whether various sensory forms, which grow out of cognitive forms, could be associated such that one would recall the other in its absence. That is, whether seeing a red object (a bottom-up facilitation), for example, would recall its weight perception to be heavier (a top-down facilitation). If so, to what extent this would be directional and transferred (e.g. whether a heavier object would recall its redness), hence share the same cognitive form.

Research in cognitive psychology investigates these associations across cultures and languages by looking at mappings between sensory forms, which are termed cross-modal correspondences [134]. Another example of cross-modal correspondences is where a higher-pitched signal of an auditory form can be associated with a higher vertical elevation of a visual form [95], and a louder sound with a brighter visual form [92]. Accordingly, sensory substitution techniques apply cross-modal associations to evoke a sensory form with another and are therefore likened to share properties of synaesthesia (for detailed reviews, see [6,115,117,153]). For the same reason, sensory substitution is considered a dual process of both top-down and bottom-up facilitations [5,50].

1.2.2 Sensory Substitution Devices as Cross-Modal Displays

Sensory substitution devices (SSDs) are essentially cross-modal displays [75] that are built, in principle, with how complementary cross-modal cues are associated with each other [106,131,134]. The mappings between visual and auditory forms, in terms of elevation and pitch, and brightness and loudness, for example, could be utilised via

SSDs to represent some features of a visual form with an auditory form [93]. In the long term, these pairings could be associated strongly such that late blind people can have visual imagery similar to that of the perception of sight via sonifications from SSDs in the absence of vision [50,105,152]. With SSDs, it is possible to acquire visual information by means of sonifications [93] or two-dimensional tactile cues [11], and auditory information by means of vibrotactile cues [23,38,46,103].

Moreover, sensory substitution techniques can be thought to transform, extend and augment our perceptual capacities by enabling a novel form of interaction with the environment [9,81,83]. For this reason, cross-modal displays, which empower sensory substitution techniques, are classified as sensory augmentation devices, whereby users could augment their sensory abilities with additional inputs such as thermal and ultrasound mapping [100]. Firefighters, for example, can sense distance via tactile gloves equipped with ultrasound sensors when their vision is restricted, thereby gaining enhanced mobility [28]. Nevertheless, sensory substitution techniques are mainly applied in the context of developing assistive technologies for the visually impaired, which, in practice, would enable them to have access to visual information via non-visual cross-modal cues [8,30,82,83,90,116].

1.2.2.1 Empirical Evidence from Applied Auditory-to-Visual and Tactile-to-Visual Sensory Substitution Techniques

In developing assistive devices, research has dominantly investigated auditory-to-visual and tactile-to-visual sensory substitution techniques. Some of these devices and their working principles will be briefly reviewed here. In the auditory domain, a vast majority of research has focused on the associations between the direction of pitch and movement (for examples of devices, see EyeMusic [1,85]; Vibe [45,62]; See ColOr [14]; The PSVA [26]; The vOICe [93]; Elektroftalm [136]; Optophone [35]). That is, the higher the elevation, the higher pitch the sonification signal has been paired with. Additional horizontal directionality is also encoded with stereoscopic and temporal properties of the sonifications that are mapped to the spatial dimension (e.g. something on the left can be heard earlier from the left earphone). Additionally,

Synaestheatre, for example, incorporated depth information via a 3D sensor and spatialised sounds such that azimuth and elevation could be conveyed spatially [59]. Once applied, these techniques were evidenced to be successful in a variety of emotion conveyance, object recognition, localisation, avoidance and navigation tasks [7,13,15–17,21,42,81,105,112, 114,121,134–137].

In the tactile domain, the cross-modal pairings are more intuitive and analogical (e.g. BrainPort [156]; Tongue Display Unit [123]; TVSS [10,11]; Optacon [87]; Optohapt [55]). That is, the sensory form of a circle, for example, can be directly conveyed on the skin (e.g. someone's back or tongue) via tactile cues presented spatially in a two-dimensional circular pattern. For enhancing navigation, tactile sensory substitution techniques particularly representing the magnetic North and providing positional information on a tactile belt or a vest were also developed and examined in depth [36,49,72,73,76,77,89,121,130,148,149]. Another line of research also investigated alternative cross-modal pairings such as conveying distance cues via the strength of vibrations (e.g. EyeCane [91]; ETA (electronic travel aid) and EOA (electronic orientation aid) [36,51,89]; UltraCane and UltraBike [128,129]). Once applied, tactile-to-visual sensory substitution techniques were also evidenced to be successful in a variety of object recognition, localisation, avoidance and navigation tasks [3,25,68,69,80,98,101,123,125, 141,156,27,30,32,42,47,51,53,56].

Many approaches have also successfully conveyed colour information via cross-modal auditory and tactile feedback (for a detailed review of SoundView, Eyeborg, Kromophone, See ColOr, ColEnViSon, EyeMusic and Creole, see [60,61]). In the recent years, a number of multisensory prototypes that utilise both auditory and tactile feedback have been further prototyped and studied in the context of spatial cognition with encouraging results (EyeCane [4,22,31,90,91]; SoV [64,71]). Overall, sensory substitution techniques have been prototyped as unisensory and multisensory display modes and successfully evidenced in a variety of use cases as assistive technologies.

1.2.2.2 Widespread Adoption Problems of Sensory Substitution Techniques

Despite their documented success in laboratory settings, sensory substitution techniques have not yet gained widespread adoption within the visually impaired population [30,82,83]. They were critiqued for their lack of generatability [82], and this was presumably resonated with the notion of "sensory substitution does not work" [8]. These arguments, however, were mainly rooted in whether sensory substitution techniques literally substitute a sensory form (i.e. 'seeing with the brain' [10], 'seeing with the skin' [155], 'seeing with sound' [94]) [82,83]. Different groups of researchers offered various explanations to why this was the case [5,6,9,40,41,82,83,132]. Recently these arguments have been constructively identified and categorised to guide future research (for a detailed review, see [30]).

The lack of widespread adoption, however, is not only the case for sensory substitution techniques. It is reported that 29.3% of assistive devices are abandoned, which has a detrimental impact on individuals with impairments, as well as the wider society [112]. The abandonment of assistive technologies is further explained with the lack of a user-centric approach, difficulty of procurement, poor performance, inability to meet the changes in user needs [112], and unaffordable financial costs [30]. While these factors successfully outline a detailed perspective on the abandonment of devices, their focus is still restrained with accessibility and assistance. We would therefore like to contribute to this framework by suggesting the adoption of an inclusive design mindset towards the applications of sensory substitution techniques, hence cross-modal displays.

As cross-modal correspondences exist across cultures and languages, and sensory substitution techniques provide sensory channel independent interactions with technological devices, they could appeal to a wider range of people, and their capabilities and needs [74]. That is, the extensive previous research in sensory substitution techniques suggests that various sensory forms (e.g. auditory or tactile) could be utilised interchangeably to have access to the same sensory information. In this way, digital interactions could switch sensory forms, and be further supplemented

and enhanced depending on user preferences and use cases. For example, sensory substitution techniques can be applied inexpensively in an open-ended sensory augmentation context [43], and therefore implemented in a variety of applications such as information and communication technologies and extended reality platforms [81,83]. With increasing interest towards inclusive cross-modal displays that appeal to larger populations, more research and development resources would be available. This could eventually overcome some of the problems that were identified with sensory substitution techniques applied as assistive technologies.

1.2.3 Potential Applications of Cross-Modal Displays for all of Us

Despite our rich multisensory capabilities within the physical world [24], relatively fewer studies have so far looked at multisensory display modes in the human-computer interactions research [135]. This, by default, not only excludes a variety of user profiles and use cases but also inhibits our multisensory capabilities from richer interactions with the digital world. From a commercial perspective, it respectively means that the competitive edge of a product is limited by a smaller market share than the potential of its inclusive alternative. Overall, disabling our capabilities via technology impacts the mobility, education, social connection and the such of a wider range of people, and intercepts the socioeconomic growth of the society as a whole.

Developing a singular product or a service that appeals to a greater number of people is challenging, if not impossible. In this sense, inclusive design does not advocate an omnipotent and omnipresent technology solution to address the unforeseen issues associated with the traditional legacy frameworks of usability and accessibility. Instead, the inclusive design mindset aims to develop flexible and modular technologies that appeal to all of us by considering our mutual perceptual and cognitive capabilities. Instead of compensating for the sensory forms of how we are able to acquire sensory information on a perceptual level, technologies can be empowered by how we process this information on a cognitive level. That is, the sensory forms we experience might be exclusive, and hence favour some groups of individuals than the others once distinguished via technology. The cognitive forms

that we are able process, on the other hand, might be inclusive, and hence connect us together.

Sensory substitution techniques, in this regard, can be applied using a supplementary rather than assistive framework [82]. They can be utilised in the context of developing artistic applications, games, extended reality environments, development of portable and intuitive systems, mobility, communication and education platforms, and interacting with novel forms of emerging information [81,83,100]. They can be used to enrich our experience with the digital world by complementing, and hence reducing, some of the visual information via non-visual cross-modal cues [65]. They can be deployed in conveying emotions via information and communication technologies, improve tangible interactions and provide navigation cues without a screen dependency via novel sensory forms. For example, a navigation application that can communicate directions via cross-modal tactile, auditory and visual forms would have the merit to be adopted by a wider range of people. All this might be possible via cross-modal displays because they would transform, extend and augment our capabilities irrespective of the sensory form [9,81]. Exceeding the sensory form would therefore bring inclusion to cross-modal displays, thereby reaching to a wider range of people. This would simultaneously improve the experiences for minority populations who were originally served by sensory substitution techniques applied as assistive technologies.

1.3 Conclusion

Inclusion is as much about technology, art, policies, social institutions, commercial models as it is about how we accept and tolerate one another in our societies. It is the mindset that can be applied in thinking, designing and creating, thereby encouraging the conversation to be in equilibrium. Overall, these premises offer an inclusive alternative to the usability and accessibility perspectives that are built on the legacy criteria of traditional frameworks, commercial models, and social and academic conversations. Instead of just converting an already existing graphical game (e.g.

Pacman or Space Invaders) into an auditory form for accessibility, for example, why not equally and collaboratively attempt to create a new form of entertainment? Why not develop new tools and approaches for novel forms of art [79,146] that can be enjoyed by a wider range of people? Why not focus on multisensory tangible interactions to democratise the "pixel empire" [68] equally with other senses? As sensory substitution stands between perception and cognition [5,50], exploring sensory substitution phenomena in this broader context could contribute new insights into how different sensory information and forms are interconnected with each other via cognitive forms. In this way, human-computer interactions can take advantage of the information processing capability of the metamodal brain in a multisensory context. Rather than creating tools which are merely assistive to compensate for the missing perceptual forms, research and development into sensory substitution techniques could be unified by a motivation for inclusion. Such cross-modal displays that are empowered with cognitive forms as much as modular and flexible sensory forms can be evaluated in a variety of use cases. Accumulated knowledge might then be transferred laterally in a multidisciplinary context, and practically applied to inclusive innovations that appeal to us all.

1.4 References

1. Sami Abboud, Shlomi Hanassy, Shelly Levy-Tzedek, Shachar Maidenbaum, and Amir Amedi. 2014. EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution. *Restorative Neurology and Neuroscience* 32, 2: 247–257. https://doi.org/10.3233/RNN-130338
2. ADA. 1990. Americans with Disabilities Act | U.S. Department of Labor. *US Public Law*, 101–336. Retrieved from https://www.dol.gov/general/topic/disability/ada
3. Junichi Akita, Takanori Komatsu, Kiyohide Ito, Tetsuo Ono, and Makoto Okamoto. 2009. CyARM: Haptic Sensing Device for Spatial Localization on Basis of Exploration by Arms. *Advances in Human-Computer Interaction* 2009: 1–6. https://doi.org/10.1155/2009/901707
4. Amir Amedi and Shlomo Hanassy. 2011. Infra red based devices for guiding blind and visually impaired persons. Retrieved from https://patents.google.com/patent/WO2012090114A1/en
5. Gabriel Arnold, Jacques Pesnot-Lerousseau, and Malika Auvray. 2017. Individual Differences in Sensory Substitution. *Multisensory Research* 30, 6: 579–600. https://doi.org/10.1163/22134808-00002561
6. Malika Auvray and Mirko Farina. 2017. Patrolling the Boundaries of Synaesthesia. In *Synaesthesia: Philosophical & Psychological Challenges*, O Deroy (ed.). Oxford University Press, Oxford, 248–274.
7. Malika Auvray, Sylvain Hanneton, and J Kevin O’Regan. 2007. Learning to Perceive with a Visuo — Auditory Substitution System: Localisation and Object Recognition with ‘The Voice.’ *Perception* 36, 3: 416–430. https://doi.org/10.1068/p5631
8. Malika Auvray and Laurence R. Harris. 2014. The State of the Art of Sensory Substitution. *Multisensory Research* 27, 5–6: 265–269. https://doi.org/10.1163/22134808-00002464
9. Malika Auvray and Erik Myin. 2009. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science* 33, 6: 1036–1058. https://doi.org/10.1111/j.1551-6709.2009.01040.x
10. Paul Bach-y-Rita, Carter C. Collins, Frank A. Saunders, Benjamin White, and Lawrence Scadden. 1969. Vision Substitution by Tactile Image Projection. *Nature* 221, 5184: 963–964. https://doi.org/10.1038/221963a0
11. Paul Bach-y-Rita and Stephen W. Kercel. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences* 7, 12: 541–546. https://doi.org/10.1016/J.TICS.2003.10.013
12. Marco Barilari, Adélaïde de Heering, Virginie Crollen, Olivier Collignon, and Roberto Bottini. 2018. Is Red Heavier Than Yellow Even for Blind? *i-Perception* 9, 1: 204166951875912. https://doi.org/10.1177/2041669518759123
13. Fernando Bermejo, Ezequiel A. Di Paolo, Mercedes X. Hüg, and Claudia Arias. 2015. Sensorimotor strategies for recognizing geometrical shapes: a comparative study with different sensory substitution devices. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00679
14. G. Bologna, B. Deville, and T. Pun. 2009. On the use of the auditory pathway to represent image scenes in real-time. *Neurocomputing* 72, 4–6: 839–849. https://doi.org/10.1016/J.NEUCOM.2008.06.020
15. Johann Borenstein. 1990. The NavBelt - A Computerized Multi-Sensor Travel Aid for Active Guidance of the Blind. In *Proceedings of the Csun’s Fifth Annual Conference on Technology and Persons with Disabilities*, 107–116. https://doi.org/10.1.1.23.9115
16. Johann Borenstein, Iwan Ulrich, and Shraga Shoval. 2000. Computerized obstacle avoidance systems for the blind and visually impaired. In *Intelligent Systems and Technologies in Rehabilitation Engineering*, H.N.L. Teodorescu and L.C Jain (eds.). CRC Press, 414–448. https://doi.org/10.1201/9781420042122.ch14
17. Assaf Botzer, Nir Shvalb, and Boaz Ben-Moshe. 2018. Using Sound Feedback to Help Blind People Navigate. In *Proceedings of the 36th European Conference on Cognitive Ergonomics - ECCE’18*, 1–3. https://doi.org/10.1145/3232078.3232083
18. John Bound and Roger Coleman. 2010. Commercial Advantage from Inclusive Design. *Design Management Review* 16, 3: 56–63. https://doi.org/10.1111/j.1948-7169.2005.tb00204.x

19. F. G. Bowe. 1991. Access to Telecommunications: The Views of Blind and Visually Impaired Adults. *Journal of Visual Impairment and Blindness* 85, 8: 328–331.
20. Julie A. Brefczynski-Lewis and James W. Lewis. 2017. Auditory object perception: A neurobiological model and prospective review. *Neuropsychologia* 105: 223–242. https://doi.org/10.1016/j.neuropsychologia.2017.04.034
21. David Brown, Tom Macpherson, and Jamie Ward. 2011. Seeing with Sound? Exploring Different Characteristics of a Visual-to-Auditory Sensory Substitution Device. *Perception* 40, 9: 1120–1135. https://doi.org/10.1068/p6952
22. Galit Buchs, Shachar Maidenbaum, and Amir Amedi. 2014. Obstacle Identification and Avoidance Using the 'EyeCane': a Tactile Sensory Substitution Device for Blind Individuals. . Springer, Berlin, Heidelberg, 96–103. https://doi.org/10.1007/978-3-662-44196-1_13
23. Austin McRae Butts. 2015. Enhancing the Perception of Speech Indexical Properties of Cochlear Implants through Sensory Substitution. Arizona State University.
24. Gemma A. Calvert, Charles Spence, and Barry E. Stein. 2004. The Handbook of Multisensory Processing.
25. Leandro Cancar, Alex Díaz, Antonio Barrientos, David Travieso, and David M. Jacobs. 2013. Tactile-Sight: A Sensory Substitution Device Based on Distance-Related Vibrotactile Flow. *International Journal of Advanced Robotic Systems* 10, 6: 272. https://doi.org/10.5772/56235
26. C. Capelle, C. Trullemans, P. Arno, and C. Veraart. 1998. A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Biomedical Engineering* 45, 10: 1279–1293. https://doi.org/10.1109/10.720206
27. Sylvain Cardin, Daniel Thalmann, and Frédéric Vexo. 2007. A wearable system for mobility improvement of visually impaired people. *The Visual Computer* 23, 2: 109–118. https://doi.org/10.1007/s00371-006-0032-4
28. Anthony Carton and Lucy E. Dunne. 2013. Tactile distance feedback for firefighters. In *Proceedings of the 4th Augmented Human International Conference on - AH '13*, 58–64. https://doi.org/10.1145/2459236.2459247
29. Mark Chamberlain, Jacqueline Esquivel, Fiona Miller, and Jeff Patmore. 2015. BT's adoption of customer centric design. *Applied Ergonomics* 46: 279–283. https://doi.org/10.1016/j.apergo.2013.03.009
30. Daniel-Robert Chebat, Vanessa Harrar, Ron Kupers, Shachar Maidenbaum, Amir Amedi, and Maurice Ptito. 2018. Sensory Substitution and the Neural Correlates of Navigation in Blindness. In *Mobility of Visually Impaired People*. Springer International Publishing, Cham, 167–200. https://doi.org/10.1007/978-3-319-54446-5_6
31. Daniel-Robert Chebat, Shachar Maidenbaum, and Amir Amedi. 2015. Navigation Using Sensory Substitution in Real and Virtual Mazes. *PLOS ONE* 10, 6: e0126307. https://doi.org/10.1371/journal.pone.0126307
32. Daniel-Robert Chebat, Fabien C. Schneider, Ron Kupers, and Maurice Ptito. 2011. Navigation with a sensory substitution device in congenitally blind individuals. *NeuroReport* 22, 7: 342–347. https://doi.org/10.1097/WNR.0b013e3283462def
33. Ilsong Choi, Jae-Yun Lee, and Seung-Hee Lee. 2018. Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology* 52: 115–122. https://doi.org/10.1016/J.CONB.2018.05.002
34. P. John Clarkson and Roger Coleman. 2015. History of inclusive design in the UK. *Applied Ergonomics* 46, PB: 235–247. https://doi.org/10.1016/j.apergo.2013.03.002
35. E. E. F. d'Albe. 1914. On a Type-Reading Optophone. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 90, 619: 373–375. https://doi.org/10.1098/rspa.1914.0061
36. D. Dakopoulos and N.G. Bourbakis. 2010. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1: 25–35. https://doi.org/10.1109/TSMCC.2009.2021255
37. Peter T. Daniels and William Bright. 1996. Analog and Digital Writing. In *The World's Writing Systems*. Oxford University Press, 886.
38. David Eagleman. 2015. Can we create new senses for humans? *TED*. Retrieved from https://www.ted.com/talks/david_eagleman_can_we_create_new_senses_for_humans
39. DDA. 1995. Disability Discrimination Act 1995. Retrieved from https://www.legislation.gov.uk/ukpga/1995/50/contents

40. Ophelia Deroy and Malika Auvray. 2012. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00457
41. Ophelia Deroy and Malika Auvray. 2014. A Crossmodal Perspective on Sensory Substitution. In *Perception and Its Modalities*. Oxford University Press, 327–349. https://doi.org/10.1093/acprof:oso/9780199832798.003.0014
42. Design Council. 2019. Design Council. *10*. Retrieved from https://www.designcouncil.org.uk/
43. Gershon Dublon and Joseph A. Paradiso. 2012. Tongueduino. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 1453. https://doi.org/10.1145/2212776.2212482
44. L. Dunai, G. Peris-Fajarnés, E. Lluna, and B. Defez. 2013. Sensory Navigation Device for Blind People. *Journal of Navigation* 66, 3: 349–362. https://doi.org/10.1017/S0373463312000574
45. Barthélémy Durette, Nicolas Louveton, David Alleysson, and Jeanny Hérault. 2008. Visuo-auditory sensory substitution for mobility assistance: testing TheVIBE. *Workshop on Computer Vision Applications for the Visually Impaired*: 1–13.
46. D. M. Eagleman, S. D. Novich, D. Goodman, A. Sahoo, and M. Perotta. 2017. Method and system for providing adjunct sensory information to a user. Retrieved from https://patents.google.com/patent/US10198076B2/en
47. Engineering Design Centre. 2019. Inclusive Design Toolkit. *University of Cambridge*. Retrieved from http://www.inclusivedesigntoolkit.com/
48. Jan B. F. van Erp, Hendrik A. H. C. Van Veen, Chris Jansen, and Trevor Dobbins. 2005. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception* 2, 2: 106–117. https://doi.org/10.1145/1060581.1060585
49. S. Ertan, C. Lee, A. Willets, H. Tan, and A. Pentland. 1998. A wearable haptic navigation guidance system. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, 164–165. https://doi.org/10.1109/ISWC.1998.729547
50. Tayfun Esenkaya and Michael J. Proulx. 2016. Crossmodal processing and sensory substitution: Is "seeing" with sound and touch a form of perception or cognition? *Behavioral and Brain Sciences* 39: e241. https://doi.org/10.1017/S0140525X1500268X
51. René Farcy, Roger Leroux, Alain Jucha, Roland Damaschinin, Colette Grégoire, and Aziz Zogaghi. 2006. Electronic Travel Aids and Electronic Orientation Aids for blind people: technical, rehabilitation and everyday life points of view. In *Conference on Assistive Technology for Vision and Hearing Impairment (CVHI)*.
52. Elise Faugloire and Laure Lejeune. 2014. Evaluation of heading performance with vibrotactile guidance: The benefits of information–movement coupling compared with spatial language. *Journal of Experimental Psychology: Applied* 20, 4: 397–410. https://doi.org/10.1037/xap0000032
53. Roy W. Feinson. 1985. Interpretive tone telecommunication method and apparatus. Retrieved from https://patents.google.com/patent/US4754474
54. Tom Froese, Marek McGann, William Bigge, Adam Spiers, and Anil K. Seth. 2012. The Enactive Torch: A New Tool for the Science of Perception. *IEEE Transactions on Haptics* 5, 4: 365–375. https://doi.org/10.1109/TOH.2011.57
55. Frank A. Geldard. 1966. Cutaneous coding of optical signals: The optohapt. *Perception & Psychophysics* 1, 11: 377–381. https://doi.org/10.3758/BF03215810
56. Asif A. Ghazanfar and Charles E. Schroeder. 2006. Is neocortex essentially multisensory? *Trends in Cognitive Sciences* 10, 6: 278–285. https://doi.org/10.1016/J.TICS.2006.04.008
57. Patricia Grant, Lindsey Spencer, Aimee Arnoldussen, Rich Hogle, Amy Nau, Janet Szlyk, Jonathan Nussdorf, Donald C. Fletcher, Keith Gordon, and William Seiple. 2016. The Functional Performance of the BrainPort V100 Device in Persons who Are Profoundly Blind. *Journal of Visual Impairment & Blindness* 110, 2: 77–88. https://doi.org/10.1177/0145482X1611000202
58. Emily Grundy. 1999. *Disability in Great Britain : results from the 1996/97 disability follow-up to the Family Resources Survey*. Dept. of Social Security by Corporate Document Services, Leeds.
59. Giles Hamilton-Fletcher, Marianna Obrist, Phil Watten, Michele Mengucci, and Jamie Ward. 2016. "I Always Wanted to See the Night Sky": Blind User preferences for Sensory Substitution Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2162–2174. https://doi.org/10.1145/2858036.2858241

60. Giles Hamilton-Fletcher and Jamie Ward. 2013. Representing colour through hearing and touch in sensory substitution devices. *Multisensory research* 26, 6: 503–32.
61. Giles Hamilton-Fletcher, Thomas D Wright, and Jamie Ward. 2016. Cross-Modal Correspondences Enhance Performance on a Colour-to-Sound Sensory Substitution Device. *Multisensory research* 29, 4–5: 337–63.
62. Sylvain Hanneton, Malika Auvray, and Barthélemy Durette. 2010. The Vibe: a versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomechanics* 7, 4: 269–276. https://doi.org/10.1080/11762322.2010.512734
63. Ian Hardy and Stuart Woodcock. 2015. Inclusive education policies: discourses of difference, diversity and deficit. *International Journal of Inclusive Education* 19, 2: 141–164. https://doi.org/10.1080/13603116.2014.908965
64. Rebekka Hoffmann, Simone Spagnol, Árni Kristjánsson, and Runar Unnthorsson. 2018. Evaluation of an Audio-haptic Sensory Substitution Device for Enhancing Spatial Awareness for the Visually Impaired. *Optometry and Vision Science* 95, 9: 757–765. https://doi.org/10.1097/OPX.0000000000001284
65. Eve Hoggan and Stephen Brewster. 2007. Designing audio and tactile crossmodal icons for mobile devices. In *Proceedings of the ninth international conference on Multimodal interfaces - ICMI '07*, 162. https://doi.org/10.1145/1322192.1322222
66. Catherine Horwill and Elli Thomas. 2019. Inclusive Design: Beyond Accessibility | Design Council. *World Architecture Magazine China - "Accessibility for All."* Retrieved from https://www.designcouncil.org.uk/news-opinion/inclusive-design-beyond-accessibility
67. Ian Hosking, Sam Waller, and P. John Clarkson. 2010. It is normal to be different: Applying inclusive design in industry. *Interacting with Computers* 22, 6: 496–501. https://doi.org/10.1016/j.intcom.2010.08.004
68. Hiroshi Ishii. 2019. SIGCHI Lifetime Research Award Talk. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, 1–4. https://doi.org/10.1145/3290607.3313769
69. Kiyohide Ito, Kiyohide Ito, Junichi Akita, Akihiro Masatani, Makoto Okamoto, Yoshiharu Fujimoto, Ryoko Otsuki, Takanori Komatsu, and Tetsuo Ono. 2012. Development of the Future Body-Finger A novel travel aid for the blind. In *Ambient 2012: The Second International Conference on Ambient Computing, Applications, Services*, 60–63.
70. Kiyohide Ito, Makoto Okamoto, Junichi Akita, Tetsuo Ono, Ikuko Gyobu, Tomohito Takagi, Takahiro Hoshi, and Yu Mishima. 2005. CyARM: an alternative aid device for blind persons. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, 1483. https://doi.org/10.1145/1056808.1056947
71. Ómar Jóhannesson, Oana Balan, Runar Unnthorsson, Alin Moldoveanu, and Árni Kristjánsson. 2016. The Sound of Vision Project: On the Feasibility of an Audio-Haptic Representation of the Environment, for the Visually Impaired. *Brain Sciences* 6, 3: 20. https://doi.org/10.3390/brainsci6030020
72. Lynette A. Jones, M. Nakamura, and B. Lockyer. 2004. Development of a tactile vest. In *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS '04. Proceedings.*, 82–89. https://doi.org/10.1109/HAPTIC.2004.1287181
73. Lynette A. Jones and Nadine B. Sarter. 2008. Tactile Displays: Guidance for Their Design and Application. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50, 1: 90–111. https://doi.org/10.1518/001872008X250638
74. J. Bern Jordan and Gregg C. Vanderheiden. 2013. Modality-Independent Interaction Framework for Cross-Disability Accessibility. In *Cross-Cultural Design. Methods, Practice, and Case Studies*. Springer, Berlin, Heidelberg, 218–227. https://doi.org/10.1007/978-3-642-39143-9_24
75. K.A. Kaczmarek, J.G. Webster, P. Bach-y-Rita, and W.J. Tompkins. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38, 1: 1–16. https://doi.org/10.1109/10.68204
76. Silke M. Kärcher, Sandra Fenzlaff, Daniela Hartmann, Saskia K. Nagel, and Peter König. 2012. Sensory Augmentation for the Blind. *Frontiers in Human Neuroscience* 6. https://doi.org/10.3389/fnhum.2012.00037

77. Kai Kaspar, Sabine König, Jessika Schwandt, and Peter König. 2014. The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and Cognition* 28: 47–63. https://doi.org/10.1016/J.CONCOG.2014.06.006
78. Simeon Keates and John Clarkson. 2004. Why the interest in inclusive design? In *Countering design exclusion*. Springer London, London, 18–33. https://doi.org/10.1007/978-1-4471-0013-3_2
79. Christine Sun Kim. 2015. Christine Sun Kim: The enchanting music of sign language | TED Talk. *TED*. Retrieved from https://www.ted.com/talks/christine_sun_kim_the_enchanting_music_of_sign_language
80. Ron Kupers, Daniel R. Chebat, Kristoffer H. Madsen, Olaf B. Paulson, and Maurice Ptito. 2010. Neural correlates of virtual route recognition in congenital blindness. *Proceedings of the National Academy of Sciences* 107, 28: 12716–12721. https://doi.org/10.1073/pnas.1006199107
81. Charles Lenay, S. Canu, and P. Villon. 1997. Technology and perception: the contribution of sensory substitution systems. In *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*, 44–53. https://doi.org/10.1109/CT.1997.617681
82. Charles Lenay and Gunnar Declerck. 2018. Technologies to Access Space Without Vision. Some Empirical Facts and Guiding Theoretical Principles. In *Mobility of Visually Impaired People*. Springer International Publishing, Cham, 53–75. https://doi.org/10.1007/978-3-319-54446-5_2
83. Charles Lenay, Olivier Gapenne, Sylvain Hanneton, Catherine Marque, and Christelle Genouëlle. 2003. SENSORY SUBSTITUTION: LIMITS AND PERSPECTIVES. In *Touching for Knowing: Cognitive Psychology of Haptic Manual Perception*, Y Hatwell, A Streri and E Gentaz (eds.). John Benjamins Publishing Company, Amsterdam, Netherlands, 275–292.
84. Shelly Levy-Tzedek, Itai Novick, Roni Arbel, Sami Abboud, Shachar Maidenbaum, Eilon Vaadia, and Amir Amedi. 2012. Cross-sensory transfer of sensory-motor information: visuomotor learning affects performance on an audiomotor task, using sensory-substitution. *Scientific Reports* 2, 1: 949. https://doi.org/10.1038/srep00949
85. Shelly Levy-Tzedek, Dar Riemer, and Amir Amedi. 2014. Color improves "visual" acuity via sound. *Frontiers in Neuroscience* 8. https://doi.org/10.3389/fnins.2014.00358
86. James W. Lewis. 2010. Audio-Visual Perception of Everyday Natural Objects – Hemodynamic Studies in Humans. In *Multisensory Object Perception in the Primate Brain*. Springer New York, New York, NY, 155–190. https://doi.org/10.1007/978-1-4419-5615-6_10
87. J.G. Linvill and J.C. Bliss. 1966. A direct translation reading aid for the blind. *Proceedings of the IEEE* 54, 1: 40–51. https://doi.org/10.1109/PROC.1966.4572
88. Dorothy Kerzner Lipsky and Alan Gartner. 1996. Inclusion, School Restructuring, and the Remaking of American Society. *Harvard Educational Review* 66, 4: 762–797. https://doi.org/10.17763/haer.66.4.3686k7x734246430
89. Jin Liu, Jingbo Liu, Luqiang Xu, and Weidong Jin. 2010. Electronic travel aids for the blind based on sensory substitution. In *2010 5th International Conference on Computer Science & Education*, 1328–1331. https://doi.org/10.1109/ICCSE.2010.5593738
90. Shachar Maidenbaum, Shelly Levy-Tzedek, Daniel-Robert Chebat, and Amir Amedi. 2013. Increasing Accessibility to the Blind of Virtual Environments, Using a Virtual Mobility Aid Based On the "EyeCane": Feasibility Study. *PLoS ONE* 8, 8: e72555. https://doi.org/10.1371/journal.pone.0072555
91. Shachar Maidenbaum, Shelly Levy-Tzedek, Daniel Robert Chebat, Rinat Namer-Furstenberg, and Amir Amedi. 2014. The Effect of Extended Sensory Range via the EyeCane Sensory Substitution Device on the Characteristics of Visionless Virtual Navigation. *Multisensory Research* 27, 5–6: 379–397. https://doi.org/10.1163/22134808-00002463
92. Lawrence E Marks. 1974. On associations of light and sound: the mediation of brightness, pitch, and loudness. *The American journal of psychology* 87, 1–2: 173–88.
93. Peter Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 2: 112–121. https://doi.org/10.1109/10.121642
94. Peter Meijer. 2019. Seeing With Sound. Retrieved from https://www.seeingwithsound.com/

95. Robert D. Melara and Thomas P. O'Brien. 1987. Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General* 116, 4: 323–336. https://doi.org/10.1037/0096-3445.116.4.323
96. Anna Mieczakowski, Sue Hessey, and P. John Clarkson. 2013. Inclusive Design and the Bottom Line: How Can Its Value Be Proven to Decision Makers? . Springer, Berlin, Heidelberg, 67–76. https://doi.org/10.1007/978-3-642-39188-0_8
97. Micah M. Murray, David J. Lewkowicz, Amir Amedi, and Mark T. Wallace. 2016. Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences* 39, 8: 567–579. https://doi.org/10.1016/J.TINS.2016.05.003
98. Saskia K Nagel, Christine Carl, Tobias Kringe, Robert Märtin, and Peter König. 2005. Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering* 2, 4: R13–R26. https://doi.org/10.1088/1741-2560/2/4/R02
99. Jeffrey E. Nash and Anedith Nash. 1982. Typing on the Phone: How the Deaf Accomplish TTY Conversations. *Sign Language Studies* 1036, 1: 193–216. https://doi.org/10.1353/sls.1982.0027
100. National Research Council. 2008. *Emerging Cognitive Neuroscience and Related Technologies*. National Academies Press, Washington, D.C. https://doi.org/10.17226/12177
101. Amy Nau, Michael Bach, and Christopher Fisher. 2013. Clinical Tests of Ultra-Low Vision Used to Evaluate Rudimentary Visual Perceptions Enabled by the BrainPort Vision Device. *Translational Vision Science & Technology* 2, 3: 1. https://doi.org/10.1167/tvst.2.3.1
102. Andrew Newman and Fiona McLean. 2004. Presumption, policy and practice. *International Journal of Cultural Policy* 10, 2: 167–181. https://doi.org/10.1080/1028663042000255790
103. Scott D. Novich and David M. Eagleman. 2014. A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In *2014 IEEE Haptics Symposium (HAPTICS)*, 1–1. https://doi.org/10.1109/HAPTICS.2014.6775558
104. Laura Ortiz-Terán, Tomás Ortiz, David L. Perez, Jose Ignacio Aragón, Ibai Diez, Alvaro Pascual-Leone, and Jorge Sepulcre. 2016. Brain Plasticity in Blind Subjects Centralizes Beyond the Modal Cortices. *Frontiers in Systems Neuroscience* 10. https://doi.org/10.3389/fnsys.2016.00061
105. Tomás Ortiz, Joaquín Poch, Juan M. Santos, Carmen Requena, Ana M. Martínez, Laura Ortiz-Terán, Agustín Turrero, Juan Barcia, Ramón Nogales, Agustín Calvo, José M. Martínez, José L. Córdoba, and Alvaro Pascual-Leone. 2011. Recruitment of Occipital Cortex during Sensory Substitution Training Linked to Subjective Experience of Seeing in People with Blindness. *PLoS ONE* 6, 8: e23264. https://doi.org/10.1371/journal.pone.0023264
106. Cesare V. Parise and Charles Spence. 2012. Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research* 220, 3–4: 319–333. https://doi.org/10.1007/s00221-012-3140-6
107. A Pascual-Leone and R Hamilton. 2001. The metamodal organization of the brain. *Progress in brain research* 134: 427–45. https://doi.org/10.1016/s0079-6123(01)34028-1
108. Achille Pasqualotto and Tayfun Esenkaya. 2016. Sensory Substitution: The Spatial Updating of Auditory Scenes "Mimics" the Spatial Updating of Visual Scenes. *Frontiers in Behavioral Neuroscience* 10. https://doi.org/10.3389/fnbeh.2016.00079
109. R. Pavlak and D. Messerschmitt. 1979. KEYPAC--A Telephone Aid for the Deaf. *IEEE Transactions on Communications* 27, 9: 1366–1371. https://doi.org/10.1109/TCOM.1979.1094561
110. Hans Persson, Henrik Åhman, Alexander Arvei Yngling, and Jan Gulliksen. 2015. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal Access in the Information Society* 14, 4: 505–526. https://doi.org/10.1007/s10209-014-0358-z
111. Charles Petzold. 1999. *Code : the hidden language of computer hardware and software*. Microsoft Press.
112. Betsy Phillips and Hongxin Zhao. 2010. Predictors of Assistive Technology Abandonment. *Assistive Technology* 5, 1: 36–45. https://doi.org/10.1080/10400435.1993.10132205
113. D. Pilling and P. Barrett. 2007. Text Communication Preferences of Deaf People in the United Kingdom. *Journal of Deaf Studies and Deaf Education* 13, 1: 92–103. https://doi.org/10.1093/deafed/enm034

114. M. R. Power, D. Power, and L. Horstmanshof. 2006. Deaf People Communicating via SMS, TTY, Relay Service, Fax, and Computers in Australia. *Journal of Deaf Studies and Deaf Education* 12, 1: 80–92. https://doi.org/10.1093/deafed/enl016
115. Michael J. Proulx. 2010. Synthetic synaesthesia and sensory substitution. *Consciousness and Cognition* 19, 1: 501–503. https://doi.org/10.1016/J.CONCOG.2009.12.005
116. Michael J. Proulx and A. Harder. 2008. Sensory substitution: Visual-to-auditory sensory substitution devices for the blind. *Tijdschrift voor Ergonomie* 6, 33.
117. Michael J Proulx and P Stoerig. 2006. Seeing sounds and tingling tongues: Qualia in synaesthesia and sensory substitution. *Anthropology & Philosophy* 7: 135–151.
118. Laurent Renier, Olivier Collignon, Colline Poirier, Dai Tranduy, Annick Vanlierde, Anne Bol, Claude Veraart, and Anne G. De Volder. 2005. Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *NeuroImage* 26, 2: 573–580. https://doi.org/10.1016/J.NEUROIMAGE.2005.01.047
119. Emiliano Ricciardi, Daniela Bonino, Silvia Pellegrini, and Pietro Pietrini. 2014. Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neuroscience & Biobehavioral Reviews* 41: 64–77. https://doi.org/10.1016/J.NEUBIOREV.2013.10.006
120. Emiliano Ricciardi and Pietro Pietrini. 2011. New light from the dark: what blindness can teach us about brain function. *Current opinion in neurology* 24, 4: 357–63. https://doi.org/10.1097/WCO.0b013e328348bdbf
121. Jennifer Rochlis. 1998. A vibrotactile display for aiding extravehicular activity (EVA) navigation in space. Massachusetts Institute of Technology.
122. Junichi Sakaki. 2016. *99*.
123. Eliana Sampaio, Stéphane Maris, and Paul Bach-y-Rita. 2001. Brain plasticity: 'visual' acuity of blind persons via the tongue. *Brain Research* 908, 2: 204–207. https://doi.org/10.1016/S0006-8993(01)02667-1
124. Douglas Schuler and Aki Namioka. 1993. *Participatory Design : Principles and Practices*. CRC Press.
125. Hervé Segond, Déborah Weiss, and Eliana Sampaio. 2005. Human Spatial Navigation via a Visuo-Tactile Sensory Substitution System. *Perception* 34, 10: 1231–1249. https://doi.org/10.1068/p3409
126. S. Shoval, J. Borenstein, and Y. Koren. 1998. Auditory guidance with the Navbelt-a computerized travel aid for the blind. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 28, 3: 459–467. https://doi.org/10.1109/5326.704589
127. Joshua H Siegle and William H Warren. 2010. Distal Attribution and Distance Perception in Sensory Substitution. *Perception* 39, 2: 208–223. https://doi.org/10.1068/p6366
128. Sound Foresight Technology. 2019. UltraBike. *Sound Foresight Technology*. Retrieved from https://www.ultracane.com/ultra_bike
129. Sound Foresight Technology. 2019. UltraCane. *Sound Foresight Technology*. Retrieved from https://www.ultracane.com/about_the_ultracane
130. Bernhard Spanlang, Jean-Marie Normand, Elias Giannopoulos, and Mel Slater. 2010. A first person avatar system with haptic feedback. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology - VRST '10*, 47. https://doi.org/10.1145/1889863.1889870
131. Charles Spence. 2011. Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics* 73, 4: 971–995. https://doi.org/10.3758/s13414-010-0073-7
132. Charles Spence. 2014. The Skin as a Medium for Sensory Substitution. *Multisensory Research* 27, 5–6: 293–312. https://doi.org/10.1163/22134808-00002452
133. Charles Spence and Cesare V. Parise. 2012. The Cognitive Neuroscience of Crossmodal Correspondences. *i-Perception* 3, 7: 410–412. https://doi.org/10.1068/i0540ic
134. Charles Spence and Cesare V. Parise. 2012. The Cognitive Neuroscience of Crossmodal Correspondences. *i-Perception* 3, 7: 410–412. https://doi.org/10.1068/i0540ic
135. Sharmila Sreetharan and Michael Schutz. 2019. Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope. *Multimodal Technologies and Interaction* 3, 1: 4. https://doi.org/10.3390/mti3010004

136. W Starkiewicz and T Kuliszewski. 1963. The 80-channel elektroftalm. In *Proceedings of the International Congress on Technology and Blindness* (2d ed.), Leslie Clark (ed.). American Foundation for the Blind, New York, 157.
137. Barry E. Stein, W. Scott Huneycutt, and M. Alex Meredith. 1988. Neurons and behavior: the same rules of multisensory integration apply. *Brain Research* 448, 2: 355–358. https://doi.org/10.1016/0006-8993(88)91276-0
138. B. Stevenson and J.L. McQuivey. 2003. *The wide range of abilities and its impact on computer technology*. Retrieved from http://www.microsoft.com/enable/research/default.aspx
139. Noelle R. B. Stiles and Shinsuke Shimojo. 2015. Auditory Sensory Substitution is Intuitive and Automatic with Texture Stimuli. *Scientific Reports* 5, 1: 15628. https://doi.org/10.1038/srep15628
140. Noelle R. B. Stiles, Yuqian Zheng, and Shinsuke Shimojo. 2015. Length and orientation constancy learning in 2-dimensions with auditory sensory substitution: the importance of self-initiated movement. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00842
141. Chloé Stoll, Richard Palluel-Germain, Vincent Fristot, Denis Pellerin, David Alleysson, and Christian Graff. 2015. Navigating from a Depth Image Converted into Sound. *Applied Bionics and Biomechanics* 2015: 1–9. https://doi.org/10.1155/2015/543492
142. Ella Striem-Amit, Laurent Cohen, Stanislas Dehaene, and Amir Amedi. 2012. Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron* 76, 3: 640–652. https://doi.org/10.1016/J.NEURON.2012.08.026
143. H. Christiaan Stronks, Ellen B. Mitchell, Amy C. Nau, and Nick Barnes. 2016. Visual task performance in the blind with the BrainPort V100 Vision Aid. *Expert Review of Medical Devices* 13, 10: 919–931. https://doi.org/10.1080/17434440.2016.1237287
144. Gary Thomas. 1997. Inclusive Schools for an Inclusive Society. *British Journal of Special Education* 24, 3: 103–107. https://doi.org/10.1111/1467-8527.00024
145. Gary Thomas, David Walker, and Julie Webb. 2006. *The Making of the Inclusive School*. Routledge. https://doi.org/10.4324/9780203135198
146. Todd Selby. 2011. *Todd Selby x Christine Sun Kim | NOWNESS*. Nowness. Retrieved from https://www.nowness.com/story/todd-selby-x-christine-sun-kim
147. UK. 2010. *Equality Act*. Statute Law Database. Retrieved from https://www.legislation.gov.uk/ukpga/2010/15/contents
148. Yon Visell. 2009. Tactile sensory substitution: Models for enaction in HCI. *Interacting with Computers* 21, 1–2: 38–53. https://doi.org/10.1016/j.intcom.2008.08.004
149. Philipp Wacker, Chat Wacharamanotham, Daniel Spelmezan, Jan Thar, David A. Sánchez, René Bohne, and Jan Borchers. 2016. VibroVision. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*, 3788–3791. https://doi.org/10.1145/2851581.2890254
150. Sam Waller, Mike Bradley, Ian Hosking, and P. John Clarkson. 2015. Making the case for inclusive design. *Applied Ergonomics* 46: 297–303. https://doi.org/10.1016/j.apergo.2013.03.012
151. Sam Waller, Pat Langdon, and John Clarkson. 2010. Designing a More Inclusive World. *Journal of Integrated Care* 18, 4: 19–25. https://doi.org/10.5042/jic.2010.0375
152. Jamie Ward and Peter Meijer. 2010. Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition* 19, 1: 492–500. https://doi.org/10.1016/J.CONCOG.2009.10.006
153. Jamie Ward and Thomas Wright. 2014. Sensory substitution as an artificially acquired synaesthesia. *Neuroscience & Biobehavioral Reviews* 41: 26–35. https://doi.org/10.1016/J.NEUBIOREV.2012.07.007
154. Jeffrey Saul Weisel. 1973. Telephonic communications system for the deaf. California State University.
155. Benjamin W. White, Frank A. Saunders, Lawrence Scadden, Paul Bach-Y-Rita, and Carter C. Collins. 1970. Seeing with the skin. *Perception & Psychophysics* 7, 1: 23–27. https://doi.org/10.3758/BF03210126
156. Wicab. 2019. Wicab, Inc. | United States | BrainPort Technologies. Retrieved from https://www.wicab.com/wicab-inc

157. Andy T. Woods, Charles Spence, Natalie Butcher, and Ophelia Deroy. 2013. Fast Lemons and Sour Boulders: Testing Crossmodal Correspondences Using an Internet-Based Testing Methodology. *i-Perception* 4, 6: 365–379. https://doi.org/10.1068/i0586
158. John S. Zelek, Sam Bromley, Daniel Asmar, and David Thompson. 2003. A Haptic Glove as a Tactile-Vision Sensory Substitution for Wayfinding. *Journal of Visual Impairment & Blindness* 97, 10: 621–632. https://doi.org/10.1177/0145482X0309701007

Reflections II

"My disability is my ability"
m-flo

The main motivation of the present thesis is to explore potential applications of sensory substitution techniques and cross-modal displays with an inclusive design mindset. Sensory substitution phenomena transform the representation of one sensory form into an equivalent form of a different sensory origin. The immediate applications of this can be seen in developing assistive technologies that aid people with sensory impairments to gain access to the information that would have otherwise been provided via the impaired sensory organ. Nevertheless, their adoption is not common.

The previous review approached the question of why sensory substitution techniques are not widely adopted from an inclusive design perspective. It further proposed that inclusion can expand their use cases, which would appeal to a wider range of people. This might be possible due to the brain's ability to process and make sense of information independent of the sensory origin (i.e. metamodal organisation of the brain). Assuming that the metamodal brain is the norm, this indicates that sensory substitution techniques can benefit everyone in the way they interact with technology. A simple example could be the way a sighted pedestrian would use a navigation application, which requires frequent screen-dependent feedback. This means of human-computer interaction will be more difficult for a pedestrian with a visual impairment. This consequently results in the development of many specialist solutions. While specialist solutions are helpful, an inclusive alternative could co-exist, which would benefit both parties equally. That is, a navigation application with cross-modal display modes would enable users to switch between sensory channels as required. This would appeal to a wide range of users.

Cross-modal displays might provide the same information via different sensory channels. A sighted pedestrian can benefit from a cross-modal auditory and/or tactile

display mode in many ways because they would substitute the frequent visual feedback from a screen. Similarly, a pedestrian with visual impairments could benefit from the same display modes. The development of cross-modal displays, therefore, unites research and development resources for technologies that can appeal to a wider group. This, in return, creates a greater business case and socioeconomic impact while explicitly breaking the arbitrary division between mainstream and assistive technologies. After all, every technology is meant to be assistive for everyone.

In the rest of the present thesis, a similar motivation will guide the reader through the exploration of potential applications of cross-modal displays for all of us. Throughout the thesis, '*display mode*' is used frequently to indicate that a cross-modal display can potentially have many display modes, such as auditory and tactile. In this way, it is meant that the same information from a cross-modal display could be conveyed via different modes. The next chapter will evaluate whether emotions could be communicated via auditory or tactile cross-modal display modes over digital communication systems.

CHAPTER II

Cross-Modal Tactile and Sonification Associations to Enrich Digital Emotion Communication

Declaration

This declaration concerns the article entitled:			
Cross-Modal Tactile and Sonification Associations to Enrich Digital Emotion Communication			
Publication status (tick one) Draft manuscript [x] Submitted [ ] In review [ ] Accepted [ ] Published [ ]			
Publication details (reference)	N/A		
Copyright status (tick the appropriate statement) I hold the copyright for this material [x] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed Formulation of ideas: 85% Design of methodology: 60% Experimental work: 70% Presentation of data in journal format: 90% For details, please see acknowledgements on the next page		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		**Date**	

Acknowledgements

I am thankful to Shantanu Bala and Troy McDaniel for their collaboration in running Experiment I. I am grateful to David Brown and Michael Proulx for their help with the analysis and discussion of the research. I am also grateful to Vanessa Lloyd-Esenkaya for patiently proof-reading the manuscript. I also thank all the participants who took part in the experiments.

2.0 Abstract

Emotions are multisensory by nature and their conveyance contributes significantly to social interactions, both 'in person' and via telecommunications. The complex nature of emotions, however, challenges their digital conveyance, especially via current unisensory computer-mediated-communications that rely heavily on visual cues. As a result, various user profiles are excluded from participation in mainstream social platforms, and multisensory capabilities of wider user groups are limited in a variety of use cases. The current study explored the feasibility of using cross-modal displays, which utilise sensory substitution techniques, to develop more inclusive affective computer-mediated communications. In a series of three experiments, the cross-modal associations between tactile, sonification and visual feedback, and basic emotional responses (happiness, surprise, anger, fear, disgust, sadness) were explored via 16 unique stimulation patterns across the US, UK and Turkey. The results showed that basic emotions could be communicated via cross-modal displays and that negative emotions could be communicated across cultures. By investigating the spatial factors that formed these stimulation patterns, the current research also aimed to formulate the foundations of a framework of cross-modal associations for designing affective cross-modal displays. The current study further demonstrated that emotion perception using cross-modal displays share characteristics of both the substituting (i.e. tactile and sonification) and substituted (i.e. visual) sensory origins. Overall, the current research discussed the findings along with limitations and future perspectives in the context of building inclusive affective technologies with cross-modal displays.

2.1 Introduction

Emotions play a fundamental role in everyday social interactions, fostering healthy personal and professional relationships. They are expressed freely, and appeal to multiple senses such as vision, hearing, touch and their various combinations [60]. Posture, gaze, gestures, facial expressions, prosody, and physical appearance all build a united multisensory context for a conversation that aids in constructing an affective interpretation of the information exchange [31]. Most of the established and mainstream computer-mediated communications (CMCs), however, rely on vision as the dominant sensory channel to communicate with the users, and the majority lacks congruent multisensory interactions. This inherently leads to various inclusivity challenges in the context of developing affective CMCs.

Visual emoticons, emoji, animated stickers, avatars or shorthand abbreviations are studied extensively in relation to whether they can enrich modern messaging systems and text-based communications, which can deliver verbal information efficiently. Such iconographic expressions of emotions, however, are found to be limited channels of information to capture the nuance of a conversation, such as expressing, receiving and perceiving contextual emotions [32]. In more complex digital social platforms (e.g. massively multiplayer online games), loading visual cues with multiple layers of contextual information (e.g. cues necessary for the gaming and social factors) is further shown to negatively influence the user experience. In these instances, the overuse of visual information inherently increases the cognitive load of users, thereby decreasing the usability of applications [63]. Unfortunately, sensory channel dependent CMCs (e.g. heavily visual ones) exclude various user profiles from participation in mainstream social platforms and prevent the multisensory capabilities of wider user groups in a variety of use cases. This consequently challenges and thereby necessitates the development of affective CMCs in an inclusive context.

To tackle these challenges, a variety of research has investigated non-visual affective information exchange. In the tactile domain, for example, using a locally linear embedding algorithm, the emotional content of facial expressions from a video feed

was mapped to a “Y” shaped vibrotactile display [49]. The tactors placed on the back of a chair provided real-time representations of three emotional dimensions (i.e. happiness, sadness and surprise). Each emotion was coded to one of the legs of the “Y” where the location of the stimulation denoted the perceived intensity of expression. This system was able to successfully deliver three basic emotions. Nevertheless, it was not flexible and modular because it could not incorporate additional emotions. An alternative series of tactile devices (i.e. HaptiHug, HaptiHeart, HaptiButterfly, HaptiShiver, HaptiTemper and HaptiTickler) were also developed to augment online and mobile messaging systems [61,62]. When actuated, these devices simulated a hug, a heartbeat, butterfly sensations, vertical axis stimulations on the spine, temperature changes, and tickling sensations respectively. Though these exemplified several other novel approaches in improving social interactions via digital communications, they did not directly map onto the emotional state of users. They were inflexible and required the use of a single tactile device for each individual feedback.

In order to develop flexible and modular affective tactile display modes, a series of research also attempted to create a tactile language [8,29,65,66]. Similar to the building blocks of language, such as the alphabet and phonemes, this line of research investigated elementary tactile stimulations. For example, various tactile stimulation patterns were given emotional meanings, which were determined either visually [8] or arbitrarily [65]. The tactile pattern-emotion pairs were then tested to validate whether they communicated basic emotions [8,64]. It was shown that the tactile language led to high accuracy in recognising basic emotions. Despite their success, the acquired tactile vocabulary was only retained in short-time memory, which limited their use in the long term [65].

Unlike the arbitrary tactile phonemes [20], it is possible to utilise ‘native’ cross-modal vocabularies. These could be thought of as the correspondences between sensory representations of different origins, which are consistent across cultures [44,52,53]. Sensory substitution devices are essentially cross-modal displays, which can deploy these correspondences [30]. In this way, they can represent one type of sensory

information via another. That is, via sensory substitution techniques, it is possible to acquire visual information, for example, by means of sonifications [38] or two-dimensional tactile cues [6], and auditory information by means of vibrotactile cues [14,41]. As taking advantage of a native language (i.e. cross-modal correspondences) is easier and more natural than acquiring a new one (i.e. an arbitrary tactile language), this consequently encourages studying cross-modal displays in delivering affective cues.

The conveyance of basic emotions is investigated in the auditory domain via cross-modal sonifications. Investigating sonifications demonstrated that blind (both congenital and late) users of a sensory substitution device, namely The vOICe [38], were able to recognise some basic emotions (happiness, surprise, and anger) from real and virtual facial expressions [59]. Other studies also revealed that the experienced users of The vOICe have shown to not only perceive the presence of another human form but also to recognise and imitate their exact body postures [58,59]. These positive results support the view that sensory substitution techniques can provide researchers with powerful tools to develop and study affective cross-modal displays. This would potentially reach to a wider range of users and enable new use cases.

If empowering the “native” cross-modal vocabulary of users can in fact enact richer and deeper social interactions via tactile and auditory feedback, this could have numerous advantages over using the symbolic representations of many arbitrary pattern-emotion associations. This can enable cross-modal displays to surpass the limitations of arbitrary “display languages”, such as the ones associated with short-term memory [65]. Additionally, since cross-modal associations link sensory information of different origins with equivalent representations, cross-modal displays can convey emotions via various display modes. That is, if a cross-modal stimulation pattern delivers a certain emotion in one sensory origin (e.g. tactile), its equivalent (e.g. via sonifications) can be easily created and expected to deliver the same emotion. As a result, this might lead to the development of inclusive affective technologies that could switch seamlessly between display modes depending on personalised user

preferences and use cases. As such an adaptive switch could be further implemented to compliment the same information with different display modes, sensory substitution techniques could also be deployed to simplify current graphical interfaces to ease the cognitive load of the users [25]. In the case of massively multiplayer online games, for example, affective cues can be isolated from visual information and instead be conveyed via non-visual cross-modal feedback.

To our knowledge, scientific research in delivering cross-modal representations via sensory substitution techniques has been limited to object recognition, localisation and navigation tasks, and colour recognition [12,13,23,24,47]. Respectively, design guidelines, heuristics and frameworks for developing affective cross-modal displays are not well established [67]. That is, the grammar rules for conveying emotions via cross-modal vocabulary are not yet established. This makes fundamental research into unimodal cross-modal paradigms and sensory substitution necessary to evaluate their applications in improving social interactions via emerging computer-mediated communication technologies. Consequently, in a series of three experiments, we explored the cross-modal associations between basic emotions, and tactile and auditory feedback along with their visual equivalents. By examining the spatial factors that formed the stimulation patterns, we also aimed to take the first steps towards establishing a framework of cross-modal associations, which could be used in designing affective display modes.

2.2 Experimental Investigation

Overall, we conceptualised two major limitations to formulating design heuristics and a framework of cross-modal associations in previous work. The first one is associated with the acquisition of a new arbitrary display language (e.g. tactile language). This becomes problematic as either a display mode distinctively serves as a single word (e.g. HaptiButterfly for butterfly sensations only [61,62]) or the display language is limited to only a few basic emotions [49]. Furthermore, these arbitrary languages are rather inconsistent with one another across devices and technologies. This makes

their acquisition difficult for the users, thereby limiting their use to short-term applications.

The second limitation concerns the cognitive steps of how users reach an emotional response. That is, via sonifications, for example, users had to recognise facial expressions first so that they could interpret the emotion associated with it later. This two-step process is rather cognitively demanding. In fact, users were required to train for 73 hours on average [59]. In spite of the flexibility and modularity of sensory substitution techniques, operationalising them primarily for recognition tasks (e.g. facial recognition prior to emotional recognition) was repeatedly shown the necessity of lengthy training requirements [33–35]. This is evidently related to the poor user experience and scarce adoption of novel cross-modal displays [16,33–35].

These limitations raise the question of whether direct cross-modal associations between basic emotions and cross-modal feedback could be established. If these associations exist, then they could be utilised to overcome the challenges associated with arbitrary display languages. They could be further applied so that users could interpret emotional responses directly without the need for recognition tasks. Overall, building a framework of cross-modal associations would not only guide the development of affective display modes but also enable a more diverse user profile receive richer sensory experiences via a variety of display modes.

In a series of two experiments, we explored the associations between basic emotions, and tactile and auditory cross-modal feedback. The emotions were selected from a mutually inclusive set of universal basic emotions proposed by various past research [19,22,43]. The basic emotions included in the current research are happiness, sadness, anger, fear, surprise and disgust. In total, 16 visually unique patterns were created and converted into cross-modal tactile and sonification feedback. Their emotional intensities were also examined.

In order to assess whether vision mediates how cross-modal correspondences are associated, in a third experiment, we studied how the visual equivalents of the cross-

modal feedback would be associated with basic emotions. Later, these associations were compared with those of tactile and sonification feedback. Specifically, we examined three possible hypotheses based on three theses of which sensory representation forms the basis of perception with cross-modal displays. From dominance thesis [10,26,46], it is argued that perception with cross-modal displays remains in the substituted sensory origin (i.e. tactile or auditory). On the basis of this assumption, it could be hypothesised that the cross-modal feedback of stimulation patterns of different origins (e.g. tactile and sonifications) would not convey the same emotion with respect to their visual equivalents (H1). Even some mutual associations were to be found, these would be rather coincidental. This would further suggest that generalised sensory substitution principles may not be used for developing a sensory origin inclusive cross-modal framework. Deference thesis claims that perception with the cross-modal displays transfers to the substituted sensory source (i.e. vision) from the substituting source (i.e. tactile or auditory) [27,40,42]. As regards this view, it could be hypothesised that associations between emotions and stimulation patterns would be identical across sensory origins (H2). This hypothesis would suggest that a singular framework of cross-modal associations can be formulated. Vertical integration thesis, on the other hand, supports that cross-modal displays can enable pre-existing capabilities of multiple senses for the given task and alter users' perception [1,3,4,17,18]. On the basis of this thesis, it could be hypothesised that while there would be some associations between emotions and stimulation patterns across sensory origins, there would also be differences (H3). Consequently, both the similarities and differences of stimulation patterns should be addressed while establishing a cross-modal framework for developing affective cross-modal displays. In order to address this and understand the associations, the spatial factors that form the stimulation patterns could be further inspected.

2.3 Methods

2.3.1 Participants

A total of 107 participants were recruited for this study: 20 (10 male) participants aged between 18 and 26 from Arizona State University, US, for Experiment I; 60 (30 male) participants aged between 18 and 45 (M = 25, S.D. = 5.8) from Sabanci University, Turkey, for Experiment II; 27 (13 male) participants aged between 19 and 59 (M = 37.60, S.D = 12.39) from University of Bath University, UK, for Experiment III. All participants self-reported having no tactile, auditory or visual impairments, and provided their informed consent prior to the onset of the experiments. Participants were briefed about the experimental procedure and the respective display modes accordingly. Anonymity of individual responses were maintained throughout the experiments. Experiment I was approved by the Institutional Review Board from Arizona State University (protocol #1308009562), and Experiment II and Experiment III were approved by University of Bath Psychology Ethics Committee (#13-204).

2.3.2 Apparatus

Two main apparatuses were utilised in the experiments. Haptic Face Display was used for Experiment I, which examined tactile feedback and emotional response. The vOICe was used for Experiment II, which examined sonification feedback in the same context. For Experiment III, which examined visual feedback, participants were only required to have a computer connected to a screen.

2.3.2.1 Haptic Face Display (HFD)

Haptic Face Display is a custom-built vibrotactile prototype, analogous to one of the early cross-modal displays, namely the Tactile Vision Sensory Substitution [5]. HFD was mounted on the back of an ergonomic mesh chair and was used to stimulate participant's backs. It was equipped with 48 3.3-volt DC eccentric rotating mass pancake shaftless vibration motors set in a 6 (row) x 8 (column) matrix. The spacing of actuators was 2cm and 4cm for horizontal and vertical dimensions respectively (Figure 2.1a). The placement of these actuators was based on a previous study, which

investigated vibrotactile spatial acuity on the torso and the back using linear arrays of vibration motors [21]. It was concluded that the optimal spacing for distinguishing neighbouring actuators when vibrated in a sequence, such as the spatiotemporal patterns used in the current research, was at least 2-3 cm [21]. Each tactor in the display was controlled by individual ATTiny88 microcontrollers connected to a common I-C bus. As each of these utilises its own timing system, spatiotemporal sequences of vibrotactile stimulation can be controlled to a precision ≤ 1ms. A tactor strip (matrix row) comprised of a single circuit board embedded with eight microcontrollers connected to eight vibrotactile actuators (Figure 2.1b). The full matrix consisted of six tactor strips connected to a control module via the I-C bus using a custom shield design for Arduino FIO. Power for the control module was from the computer via USB. Stimulation parameters were transferred via a virtual serial port.

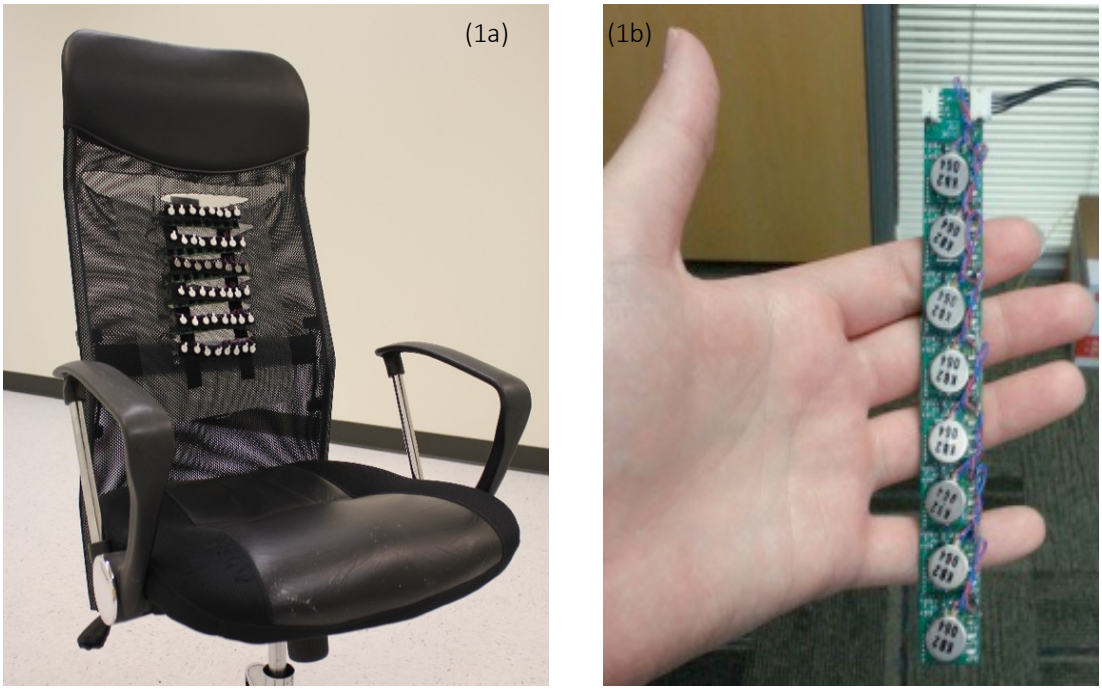


Figure 2.1a (left) displays the Haptic Face Display mounted on the back of an ergonomic mesh chair. **Figure 2.1b** (right) show a single tactor strip embedded on the HFD.

Due to hardware limitations, such as the size of custom printed circuit boards and tactor strips, six single tactor strips with eight actuators each were built and placed on HFD for reliability. This inherently limited the design options for the stimulation patterns. A 6x8 matrix structure was inevitably chosen by researchers (see *Stimuli*

section for further details) for its potential to utilize numerous distinct configurations of additional stimulation patterns.

A custom designed software interface using a tactile display driver written in Python was used to ensure consistent collection of user responses. An HTTP web server, capable of sending vibrotactile pattern commands to the control module, was created using the Flask web application framework. Using the server's HTTP interface, a web site was rendered to the user using the Jinja2 templating engine for Python. Its interface was implemented with Bootstrap and jQuery UI modules to provide a direct feedback in the Google Chrome web browser. This web page updated automatically as new vibrotactile patterns were presented on the tactile display. The chair was connected via USB to a 15-inch Lenovo Z580 laptop running Windows 7 with the custom web site open in Google Chrome.

2.3.2.2 The vOICe

The vOICe is a commercially available (freely at [39]) cross-modal display that can represent visual images via sonifications. The cross-modal algorithm of The vOICe maps visual pixels with respect to visual and auditory cross-modal correspondences, and produces a temporal audio signal to deliver additional left-to-right directionality via stereo headphones [38]. This encoding is completed by representing higher elevation with higher pitch, and brighter pixels with louder auditory signals.

2.3.3 Stimuli

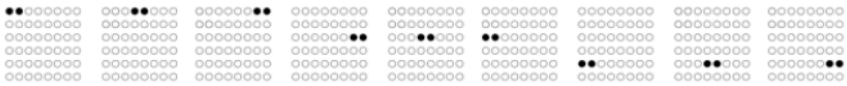
2.3.3.1 Stimulation Patterns

The creation of stimulation patterns was inspired by previous research [36], existing patterns used in affective tactile display modes [8], others used for tactile applications such as navigation [45] and common wisdom such as “a chill down the spine”. The large pool of stimulation patterns was further narrowed down with pilot testing. This process involved 20 additional participants and utilised HFD to assess the ease of

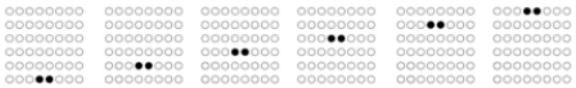
recognition of stimulation patterns. Results revealed consensus among 16 unique patterns, which were then used in the current research (Figure 2.2).

Having asked five additional coders to classify the stimulation patterns from their visual animations, six spatial factors were identified based on the descriptions of the coders. These factors were static/dynamic, scattered/clustered, leftwards/rightwards, downwards/upwards, horizontal/vertical, and alternating/sequential. In respect to these spatial factors, the stimulation patterns were plotted as polar maps (Appendix 2.A). These factors displayed how a stimulation pattern moved across the sensory space.

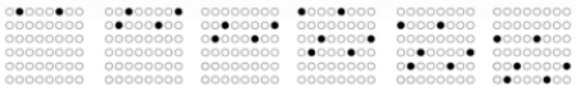
“*Horizontal Snake*” pattern. A vertical variation was also used:



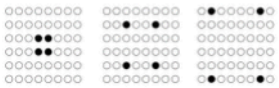
‘*Spine Up*’ pattern. A downward variation was also used:



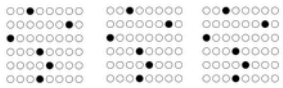
‘*Rain*’ pattern:



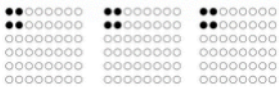
‘*Explode*’ pattern:



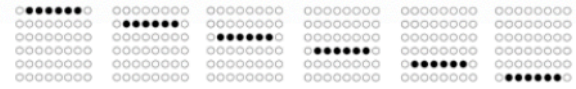
‘*Six Motor Burst*’ pattern:



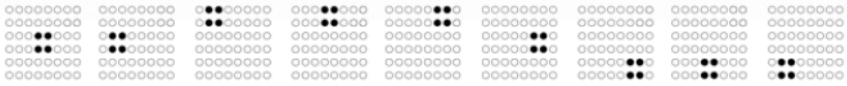
‘*Shoulder Tap*’ pattern:



‘*Wave Down*’ pattern. Up, left, and rightward variations were also used:



‘*Spiral Out*’ pattern. An inward variation was also used:



‘*Alternate Top-Bottom*’ pattern. A left-right variation was also used:

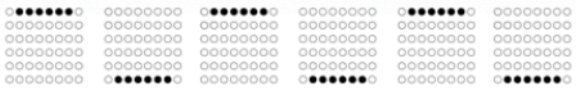


Figure 2.2 represents the stimulation patterns frame by frame. They are named arbitrarily for convenience. Each frame represents part of a linear sequence of the stimulation patterns from left to right and is presented for 500ms via each display mode.

2.3.4 Presentation of the Feedback

2.3.4.1 Experiment I: Tested Associations between Cross-Modal Tactile Feedback and Basic Emotions

HFD was programmed to simultaneously stimulate a specific set of actuators, with equal intensity, that topographically corresponded to the black dots in a stimulation frame (6x8). These frames were presented to participants in the given order (Figure 2.2). Prior to the Experiment I, participants were advised to wear a single layer of clothing to reduce dampening of vibrations produced by HFD. On the experiment day, they were seated on the HFD mounted office chair and asked to make themselves comfortable. Participants wore sound isolating headphones playing soft white noise to prevent any external noises or the subtle buzz of the vibration motors from influencing the responses during the experiment.

2.3.4.2 Experiment II: Tested Associations between Cross-Modal Sonification Feedback and Basic Emotions

The visual representations used for sonifications graphically differed from what is presented in Figure 2.2. The empty circles were eliminated and then the colours were reversed in each stimulation frame. Sonifications were created analogous to the tactile feedback, where each frame was sonified for 500ms at the default settings of The vOICe. That is, each white dot was sonified with equal intensity while the black background was silent. Each sonified frame was later sequentially combined in Audacity [2]. Similar to the tactile feedback, the sonification feedback from each stimulation pattern differed in duration due to the different numbers of frames per pattern.

2.3.4.3 Experiment III: Tested Associations between Visual Feedback and Basic Emotions

The visual feedback for each stimulation pattern were created by sequentially joining individual stimulation frames using a video editor, resulting in an animation of the successive frames where each frame was visible for 500ms. The empty circles were also eliminated during the creation of the visual feedback.

Sonification and visual feedback were uploaded to Qualtrics database [48]. For Experiment II, participants were explicitly instructed to wear stereo headphones on the correct way around and to set the volume at a comfortable level. Experiment III did not include any audio components and participants were asked to complete the study in a quiet area.

2.3.5 Experiment

Participants' responses to Experiment I, Experiment II and Experiment III were collected at independent sites respectively in USA, Turkey and the UK. While the data collection for Experiment I was lab based due to experimental requirements (i.e. the use of HFD), self-explanatory online experiments were set separately for Experiment II and Experiment III. The experiments were completed in English. Considering the universality of basic emotions and cross-modal correspondences, a cross-cultural between-subject experimental design was considered appropriate for the purposes of the overall study. Results therefore would offer both independent analyses of each experiment and a general evaluation.

2.3.5.1 Experimental Procedure

The three experiments followed the exact experimental procedure where all 16 stimulation patterns were presented in a random order via the display mode specific to the experiment. In Experiment I, each stimulation pattern was repeated three times. In Experiment II and Experiment III, the presentation of stimulation patterns was not limited by a view count; however, metadata from the online experiments indicated that participants repeated a stimulation pattern two or three times on average.

Once a stimulation pattern was delivered, participants were asked "What do you feel the pattern represents?". A fixed set of basic emotions were displayed on the computer in front of them for ease of response. The choices presented in a random order were 'Happiness', 'Sadness', 'Anger', 'Fear', 'Surprise', 'Disgust'. 'Neutral' was

also included in this set of options for Experiment I, and 'Relaxed' in Experiment II and Experiment III. This was done to test whether stimulation patterns led to a different emotional response than neutral/relaxed. Following this question, participants were then asked to rank the intensity of the conveyed emotions on a 5-point Likert scale for each stimulation pattern, where higher scores indicated higher elicitation or vice versa.

2.4 Results

The primary objective of the current research was to explore associations between basic emotions and stimulation patterns of cross-modal tactile, sonification and visual origins. While it was expected that stimulation patterns might be biased towards conveying specific emotional responses, taking the three hypotheses into consideration, the consistency of associations across the display modes were inspected. Therefore, the associations and their emotional intensities were analysed independently prior to an overall evaluation. All the analyses were completed with SPSS25 [28].

2.4.1 Strong Associations between Emotional Responses and Stimulation Patterns

The frequencies of 16 stimulation patterns were cross tabulated with respect to emotion types to form a contingency table (Appendix 2.B). These frequencies were not equal, suggesting that emotions were associated with some combination of stimulation patterns and not others. Correspondence analysis, calculated for each column of this contingency table, generated significant results for tactile ($\chi^2(90) = 243.978$, $p<.0005$), sonification ($\chi^2(90) = 515.217$, $p<.0005$) and visual ($\chi^2(90) = 213.017$, $p<.0005$) feedback. This implied strong interdependency between emotional responses and stimulation patterns. Consequently, the percentages of stimulation patterns yielding strong associations with each emotional response were examined. The probability of having a pattern associated with an emotion by random chance is ~6% in an even distribution (95% CI [.02, .12]). Associations with 12% of total

responses or higher were therefore considered to be strong associations and reported in Figure 2.3a.

Following the analysis of contingency tables via chi-square statistics, correspondence analysis was considered to be appropriate to determine the proximal relationships between emotional responses and stimulation patterns and visualise these in a low-dimensional space as biplots. In these visualisations, the distances between variables would further reveal how similar they are to each other. This can be achieved in SPSS with correspondence analysis under dimension reduction and the results can be plotted as biplots. Similar plots for the current study were created using 2 dimensions, with RcMean standardization, symmetrical normalisation and chi square statistics (Figure 2.3b). In the biplots, emotional response frequencies specific to the display mode were standardised to sum to 1.0 and represented in terms of the distance between individual rows and/or columns in low-dimensional space. For example, by inspecting the biplots, it could be seen that *wave left* stimulation pattern was strongly associated with 'anger' via tactile and sonification display modes, and *explode* with 'surprise' via the visual feedback.

2.4.2 Emotional Intensity

2.4.2.1 Within Tactile Feedback

An ANOVA analysis resulted in an omnibus main effect ($F(6,908) = 5.661$, $p<.0005$, $\eta_p^2 = 0.36$), demonstrating that the intensity of emotional responses significantly varied with respect to the emotion type. Bonferroni corrected planned contrasts further illustrated that this was mainly due to the negative emotions 'anger' and 'fear'. For the former, contrasts showed a significantly higher intensity of emotional response compared to 'sadness' (MD = 0.470, 95% CI [0.009, 0.930], $p = .041$), 'disgust' (MD = 0.541, 95% CI [0.129, 0.954], $p = .001$), and 'neutral' (MD = 0.520, 95% CI [0.069, 0.972], $p = .010$). 'Fear' also evoked a higher intensity compared to 'disgust' (MD = 0.520, 95% CI [0.113, 0.927], $p = .002$), and 'neutral' (MD = 0.498, 95% CI [0.052, 0.945], $p = .015$).

2.4.2.2 Within Sonification Feedback

A second ANOVA analysis also resulted in an omnibus main effect ($F(6,953) = 3.349$, $p = .003$, $\eta_p^2 = 0.21$) among the intensities of emotional responses to the sonification feedback. Planned contrasts further revealed that 'fear' significantly evoked a higher intensity than 'sadness' (MD = 0.465, 95% CI [0.002, 0.929], $p = .048$, and 'Relaxed' (MD = 0.540, 95% CI [0.111, 0.968], $p = .003$).

(3a)

	Tactile (%)	Sonification (%)	Visual (%)
HAPPINESS (18.6%)			
Explode	4.6	15.3	9.3
Rain	2.1	15.7	8
Spine Up	13.9	6.9	8
Vertical Snake	12.4	6.3	9.3
Wave Up	5.2	6.9	13.3
ANGER (13.0%)			
Alternate Left-Right	12	11.5	15.3
Alternate Top-Bottom	3.8	5	15.2
Wave Left	12.8	17.4	6.5
Wave Right	12	20.7	8.7
SADNESS (10.2%)			
Spine Down	6.7	25.8	22.9
Wave Down	7.9	25.8	18.8
SURPRISE (16.9%)			
Alternate Top-Bottom	15.3	6.8	10.8
Explode	7.6	9.6	16.9
Rain	6.1	6.8	13.5
Six Motor Burst	7.6	12	8.4
FEAR (16.8%)			
Alternate Left-Right	16.4	13.9	10.7
Alternate Top-Bottom	7.1	14.9	7.8
Wave Left	6.4	9.8	12.5
DISGUST (11.5%)			
Horizontal Snake	10.9	4.7	15.7
NEUTRAL (10.4%) **/RELAXED** (16.3%)			
Spiral Out	5.3	8	12.2
Vertical Snake	4.2	13.6	6.7

(3b)

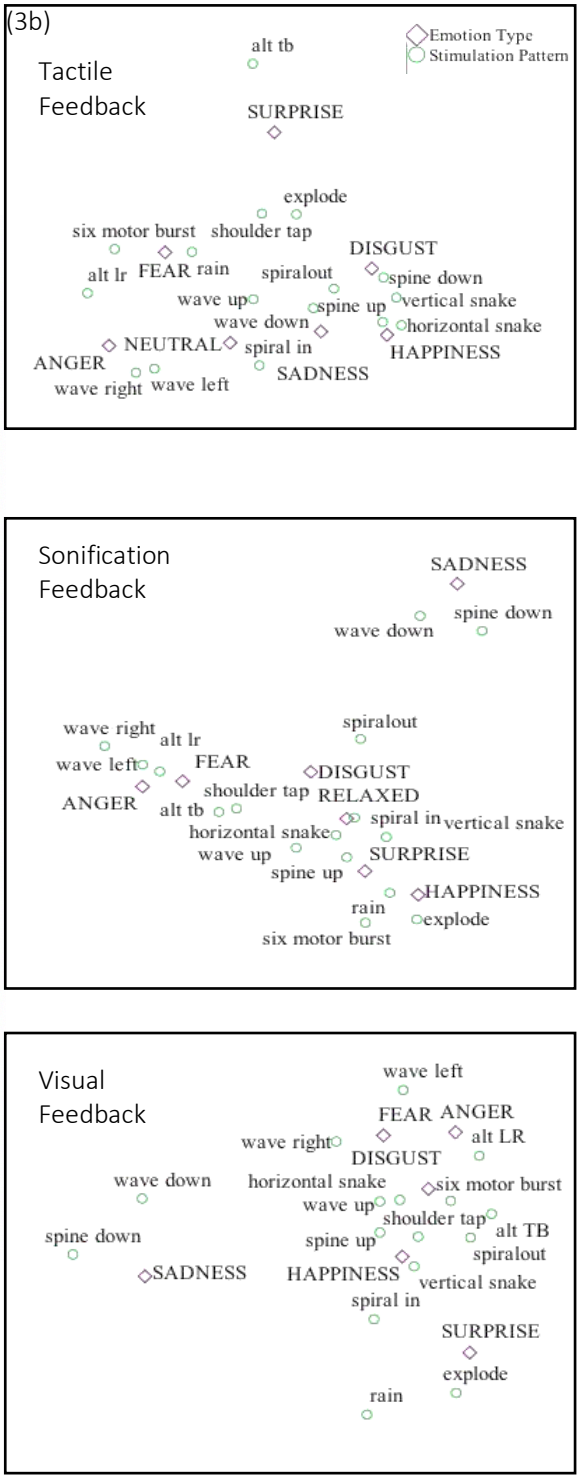


Figure 2.3a (left) shows the contingency table of strong associations between emotional responses and stimulation patterns with respect to each display mode (tactile, sonification, visual). These frequencies are provided in percentages of total responses given to a stimulation pattern associated with an emotional response. Lightly shaded cells indicate strong associations. Additionally, percentages of overall emotional response frequencies are presented next to the emotional response. Figure 2.3b (right) displays the biplots representing associations between stimulation patterns (green) and emotion types (purple) with respect to each display mode. Some stimulation pattern names are shortened (e.g. 'Alternate Top-Bottom' is 'alt tb' for aesthetics reasons).

2.4.2.3 Within Visual Feedback

Another ANOVA analysis demonstrated a non-significant main effect ($F(6, 432) = 1.186$, $p = .313$, $\eta_p^2 = 0.16$) with no significant planned contrasts between emotional responses and their intensities.

2.4.2.4 Overall Comparisons

Regardless of the emotion type, an ANOVA analysis was carried out between overall emotional intensities from the tactile, sonification and visual feedback to assess whether one display mode resulted in stronger emotional intensity (Table 2.1). The analysis showed a main overall effect, $F(2,2304) = 202.698$, $p<.0005$, $\eta_p^2 = 0.15$. Bonferroni corrected contrasts further demonstrated that the highest intensity was attributed to tactile display mode (M = 3.15, SD = 1.12) being higher than both sonification (MD = 1.082, 95% CI [0.948, 1.215], $p<.0005$) and visual (MD = 0.899, 95% CI [0.731, 1.068], $p<.0005$) feedback. Intensity from visual feedback (M = 2.25, SD = 1.37) was also significantly higher than the sonification feedback (M = 2.07, SD = 1.21) with a mean difference of 0.182, ([0.015, 0.350], $p = .028$).

Table 2.1 summarises the average (∓ SD) of emotional intensities with respect to display mode and emotion type.

	Tactile (Mean∓ SD)	Sonification (Mean∓ SD)	Visual (Mean∓ SD)	Overall (Mean∓ SD)
HAPPINESS	3.06 ∓ 1.12	2.17 ∓ 1.21	2.42 ∓ 1.35	2.62 ∓ 1.26
ANGER	3.44 ∓ 1.07	2.11 ∓ 1.19	2.21 ∓ 1.11	2.71 ∓ 1.3
SADNESS	2.97 ∓ 1.08	1.88 ∓ 1.19	1.94 ∓ 1.34	2.31 ∓ 1.29
SURPRISE	3.28 ∓ 1.1	2.07 ∓ 1.15	2.22 ∓ 1.24	2.51 ∓ 1.28
FEAR	3.42 ∓ 1.2	2.35 ∓ 1.15	2.42 ∓ 1.37	2.75 ∓ 1.31
DISGUST	2.89 ∓ 1.05	1.94 ∓ 1.17	2.59 ∓ 1.47	2.48 ∓ 1.23
RELAXED/ NEUTRAL	2.92 ∓ 1.06	1.81 ∓ 1.37	2.12 ∓ 1.56	2.24 ∓ 1.42
Overall	3.15 ∓ 1.12	2.07 ∓ 1.21	2.25 ∓ 1.37	2.53 ∓ 1.31

2.5 Discussion

The results demonstrated strong associations between basic emotions and 14 of the stimulation patterns, via cross-modal tactile and sonification feedback, and their visual equivalents. Two of the stimulation patterns (i.e. *shoulder tap* and *spiral in*) did not result in any strong associations and eight unique patterns only conveyed one type of emotion. While the associations were predominantly specific to the display mode for positive emotions ('happiness' and 'surprise'), a cross-modal consensus for the negative emotions ('anger', 'sadness' and 'fear' with the exception of 'disgust') was observed between tactile and sonification, tactile and visual, and visual and sonification feedback. That is, the same stimulation patterns resulted in the same negative emotional responses via visual, tactile and sonification display modes. Only *alternate top-bottom* was excluded from this as it was chosen for 'anger' via visual feedback, 'surprise' via tactile feedback and 'fear' via sonification feedback. This finding contributes to the previous research that investigated cross-cultural recognition of basic emotions. It was shown that, via vocalisations, negative emotions could be recognised across cultures while positive emotions were communicated with culture-specific cues [50]. Moreover, the tactile feedback significantly conveyed the strongest emotional intensity, and sonification feedback significantly resulted in the weakest. There was also a variation in emotional intensities within display modes. While these differences were not significant via the visual display mode, 'anger' and 'fear' mainly caused significantly stronger emotional intensity than other negative emotions via the tactile and sonification feedback.

The aforementioned cross-modal consensus might be intuitively deduced from vision's role as a proxy in sensory substitution techniques that empower cross-modal displays (H2). Deference thesis forms the basis of this hypothesis by arguing that perception with the cross-modal displays transfers to the substituted sensory source (i.e. vision) from the substituting source (i.e. tactile or auditory) [27,40,42]. That is, the tactile and sonification feedback were derived from visual representations, and such cross-modal feedback could be reconstructed back as visual. Consequently, it would be expected that the tactile and sonification feedback, and their visual equivalents

would frequently lead to the conveyance of the same emotional responses (H2). The results conversely suggested the opposite as there were no instances of a stimulation pattern conveying the same emotion type via the three display modes. Moreover, there were instances of stimulation patterns conveying the same emotion type via tactile and sonification feedback, excluding the visual feedback. These suggested that vision did not meditate the emotional perception with cross-modal displays.

In contrast, according to the dominance thesis, it could have been that perception with cross-modal displays would remain in their sensory origin [10,26,46]. This would further suggest that mutual associations between emotional responses and display modes would be rather coincidental, if there were any (H1). This was also found to be an unlikely explanation as cross-modal consensus was repeatedly observed between pairs of display modes, especially within the negative basic emotions.

Vertical integration thesis suggests that both the substituting and substituted sensory channels enable pre-existing capabilities of multiple senses for the given task, such as emotion conveyance, and can alter users' perception [1,3,4,17,18]. In other words, vertical integration thesis argues that the basis of perception with cross-modal displays is on a spectrum between the substituting and substituted senses, carrying the characteristics of both at varying degrees. Respectively, it was hypothesised that there would be some associations between emotion and stimulation patterns across sensory origins (H3). The results revealed there were strong associations within negative emotions but not within the positive ones. This further supported the vertical integration thesis. As positive emotions are thought to form social cohesion within a community, it is argued that cues that convey positive basic emotions, unlike negative ones, might further evolve to be restricted to in-group members of a community [50,51]. This might prevent cross-modal associations for positive emotions from existing cross-culturally. Accordingly, with reference to our cross-cultural participant pool from the US, UK and Turkey, we did not find any mutual strong associations between positive emotions and display modes that delivered the same stimulation pattern. The current research thus extends the conveyance of basic emotions to cross-modal displays. It demonstrates how cross-modal feedback and sensory substitution

techniques could successfully convey positive and negative emotions. This implies that establishing a framework of cross-modal associations should account for both the similarities and dissimilarities in how stimulation patterns lead to emotional responses with respect to display modes (H3).

2.5.1 Spatial Cross-modal Associations with Basic Emotions

Cross-modal displays can mutually convert the spatial properties of visual information into cross-modal tactile and sonification feedback. This raises the question of whether there was a "spatial rule" that governed how stimulation patterns were associated with basic emotions via the three display modes. In order to investigate this further, six spatial factors (static/dynamic, scattered/clustered, leftwards/rightwards, downwards/upwards, horizontal/vertical, and alternating/sequential) were identified with respect to how stimulation patterns moved across the sensory space (see *Stimuli* section for further details). These factors were plotted to visualise each stimulation pattern as a polar map (Appendix 2.A). Consequently, by overlapping the polar maps of stimulation patterns that conveyed the same emotional response, it might be possible to examine the spatial factors that influenced the conveyance of emotional responses (Figure 2.4). For example, by studying 'happiness' in Figure 2.4, it could be argued that the patterns that were delivered with visual, tactile and auditory displays modes were all dynamic. Pinpointing the inclusive and exclusive spatial factors with respect to the display modes might further build a cross-modal framework of how different patterns can be created to communicate emotions.

Figure 2.4 displays the polar maps of emotional responses, which were created by overlapping the polar map of each stimulation pattern that was strongly associated with the given emotion. Green, orange and grey respectively represents sonification, tactile and visual display modes. Darker connections indicate areas where there are more overlaps.

2.5.2 Establishing a Framework of Cross-Modal Associations, Limitations and Future Perspectives

As a first step towards establishing a framework, we tentatively examined the intersection points of spatial factors represented by the overlapping polar maps (Figure 2.4). For example, among the stimulation patterns that communicated 'happiness', dynamic and vertical movements were mutually inclusive to tactile, sonification, and visual feedback. For 'anger', these joint features were dynamic and clustered. For 'surprise', a greater number of factors (i.e. dynamic, downwards, vertical) were shared between tactile and sonification feedback. On the other hand, some factors, such as sequential in 'fear' communication, was exclusive to the visual feedback. Moreover, the presence or absence of a spatial factor (e.g. leftwards and rightwards) did not always influence the emotional response (e.g. 'anger' via tactile and sonification feedback).

Overall, inspecting the overlaps, or the lack thereof, might further reveal spatial rules that help establish a framework of cross-modal associations. This attempt would subsequently be an analogue to developing a grammar for native display languages to communicate emotions. That is, if stimulation patterns that elicited special emotional responses could be reduced to core spatial factors, the remaining factors could be flexibly utilised to expand the repertoire of affective display vocabulary. This approach might eventually overcome the current limitations of arbitrary display languages. The current research also suggests that it is the combination of spatial factors, rather than singular categories, that drives cross-modal associations between emotions and stimulation patterns. Future research should therefore robustly evaluate various combinations of stimulation factors, which match one to one with stimulation patterns, and their associations with emotions.

It was evident that some emotional responses were only conveyed by one (e.g. 'disgust' via visual feedback) or two (e.g. 'relaxed' and 'sadness' via sonification and visual feedback) display modes. Despite being superior in conveying the strongest emotional intensity, the tactile feedback was not successful in delivering emotions

such as sadness and disgust. Previous research has shown, however, that 'sadness' and 'disgust' can be recognised via tactile display modes once the associations between emotions and stimulation patterns are learnt [37]. This indicates that cross-modal associations might not be available for conveying an inclusive list of emotions in every sensory origin (e.g. 'sadness' and 'disgust' via cross-modal tactile feedback). It is, however, possible to complement the missing emotional traits via means of other cross-modal feedback (e.g. 'sadness' via sonification and visual feedback). The immediate implication of this resembles how most iconographic expressions of emotions, that are superiorly visual and are learnt to be associated, are not able to capture the nuance of a conversation [32]. In these instances, complementary cross-modal feedback might be utilised to enhance the emotion communication for richer user experiences.

Complementary use of cross-modal display modes further raises the question of what happens when stimulation patterns (either the same or different) are delivered in a multisensory context to enrich the user experience [15,54]. It could be hypothesised that the multisensory feedback might enhance (hence additive) or fade out (hence subtractive) the overall intensity of emotional responses [55–57]. The results of the current study indicated that the tactile display mode resulted in the strongest intensity while the sonification feedback was the weakest. However, this does not reveal, for example, the emotion intensity of a sonification-tactile feedback. When primary colours overlap, the sum of their parts creates an additional colour. Similarly then, it could be hypothesised that a multisensory response to stimulation patterns would communicate emotions beyond their basic subset.

It is important to highlight that discrete emotional models may not explain the affective experiences of all individuals [7]. The current research is limited to a mutually inclusive set of universal basic emotions proposed by various past research [19,22,43]. Utilising universal emotions was necessary to explore the cross-modal associations between stimulation patterns and emotions across cultures. Furthermore, the use of discrete emotions allows one to focus on the cross-modal associations with minimised cognitive load of the experimental tasks. Other assessments based on dimensional

emotion models, such as self-assessment manikin [11], and valence and arousal [7] were therefore not included. While dimensional emotion assessments were not appropriate for the current study, it is possible that some of the participants responded to the affective feedback in a way which could not be captured by the methods employed here. The prior emotional state of participants was not evaluated either. This design decision was taken as it is equally important for affective displays to communicate emotions consistently in all scenarios. In respect of this, the high statistical power obtained in this study due to the large sample size (N = 107) reduced the risk of such influences skewing the overall results. Despite these variables, this research demonstrated that cross-modal displays could communicate basic emotions and especially negative ones across cultures via the same stimulation patterns.

It is, however, important to note that the statistical results reported here are a result of nonparametric correspondence analysis, which utilised multiple tests. In total, there were 27 strong associations between emotional responses and stimulation patterns across tactile, sonification and visual cross-modal feedback. Using an alpha value of .05, it can be assumed that 0.05 analyses will be significant by chance. Therefore, 1.4 out of the 27 associations in the correspondence analyses shown here could be significant due to chance alone. The study presented here is exploratory and does not present hypotheses on which emotional responses and stimulation patterns would be paired. Adjustments for multiple tests might therefore be impossible without a prespecified hypothesis [9]. Future research might expand the framework of cross-modal associations between emotional responses and stimulation patterns using effect sizes as well as looking at statistical significance. In this respect, the framework provided here could be used for future hypothesis testing allowing the necessary statistical adjustments to be taken.

2.6 Conclusion

The current research has successfully demonstrated how tactile, sonification, and visual feedback could be studied to communicate basic emotions via cross-modal displays. Overall, 14 out of the 16 unique stimulation patterns utilised in the current research were strongly associated with basic emotions. A cross-modal consensus between tactile and sonification, tactile and visual, and visual and sonification feedback were found, especially among negative emotions across cultures. This extends the previous research, which showed that the expressions of negative emotions via vocalisations were shared cross-culturally, to cross-modal displays. The findings further support the vertical integration thesis, which argues that perception with cross-modal displays enables the capabilities of both the substituting and substituted sensory origins. This suggests that the similarities and differences in cross-modal associations between distinct sensory origins should be considered while establishing a cross-modal framework for developing affective cross-modal displays. Consequently, it is tentatively suggested that investigating the inclusive and exclusive spatial factors of stimulation patterns via tactile, sonification and visual feedback that conveyed the same emotional response might formulate a framework of cross-modal associations. The current framework is still in an early phase and requires replication across other bodies of research to generalise the spatial cross-modal associations between emotions and stimulation patterns. Once matured, it could be operationalised in developing inclusive affective display modes. Overall, the current research argues that studying cross-modal displays is a good candidate for inclusion as they could provide the same information via various forms of display modes depending on the user preferences and use cases.

2.7 References

1. Gabriel Arnold, Jacques Pesnot-Lerousseau, and Malika Auvray. 2017. Individual Differences in Sensory Substitution. *Multisensory Research* 30, 6: 579–600. https://doi.org/10.1163/22134808-00002561
2. Audacity. 2019. Audacity. *Audocity*. Retrieved from https://www.audacityteam.org/
3. Malika Auvray and Mirko Farina. 2017. Patrolling the Boundaries of Synaesthesia. In *Synaesthesia: Philosophical & Psychological Challenges*, O Deroy (ed.). Oxford University Press, Oxford, 248–274.
4. Malika Auvray and Erik Myin. 2009. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science* 33, 6: 1036–1058. https://doi.org/10.1111/j.1551-6709.2009.01040.x
5. Paul Bach-y-Rita, Carter C. Collins, Frank A. Saunders, Benjamin White, and Lawrence Scadden. 1969. Vision Substitution by Tactile Image Projection. *Nature* 221, 5184: 963–964. https://doi.org/10.1038/221963a0
6. Paul Bach-y-Rita and Stephen W. Kercel. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences* 7, 12: 541–546. https://doi.org/10.1016/J.TICS.2003.10.013
7. Lisa Feldman Barrett. 1998. Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition and Emotion* 12, 4: 579–599. https://doi.org/10.1080/026999398379574
8. M. Benali-Khoudja, M. Hafez, A. Sautour, and S. Jumpertz. 2005. Towards a new tactile language to communicate emotions. In *IEEE International Conference Mechatronics and Automation, 2005*, 286–291. https://doi.org/10.1109/ICMA.2005.1626561
9. R. Bender and S. Lange. 1999. Multiple test procedures other than Bonferroni's deserve wider use [5]. *British Medical Journal 318*, 600–601. https://doi.org/10.1136/bmj.318.7183.600a
10. Ned Block. 2007. Spatial Perception via Tactile Sensation. In *Consciousness, Function, and Representation*. The MIT Press. https://doi.org/10.7551/mitpress/2111.003.0020
11. Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1: 49–59. https://doi.org/10.1016/0005-7916(94)90063-9
12. David J. Brown, Andrew J. R. Simpson, and Michael J. Proulx. 2014. Visual Objects in the Auditory System in Sensory Substitution: How Much Information Do We Need? *Multisensory Research* 27, 5–6: 337–357. https://doi.org/10.1163/22134808-00002462
13. David Brown, Tom Macpherson, and Jamie Ward. 2011. Seeing with Sound? Exploring Different Characteristics of a Visual-to-Auditory Sensory Substitution Device. *Perception* 40, 9: 1120–1135. https://doi.org/10.1068/p6952
14. Austin McRae Butts. 2015. Enhancing the Perception of Speech Indexical Properties of Cochlear Implants through Sensory Substitution. Arizona State University.
15. Gemma A. Calvert, Charles Spence, and Barry E. Stein. 2004. The Handbook of Multisensory Processing.
16. Daniel-Robert Chebat, Vanessa Harrar, Ron Kupers, Shachar Maidenbaum, Amir Amedi, and Maurice Ptito. 2018. Sensory Substitution and the Neural Correlates of Navigation in Blindness. In *Mobility of Visually Impaired People*. Springer International Publishing, Cham, 167–200. https://doi.org/10.1007/978-3-319-54446-5_6
17. Ophelia Deroy and Malika Auvray. 2012. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00457
18. Ophelia Deroy and Malika Auvray. 2014. A Crossmodal Perspective on Sensory Substitution. In *Perception and Its Modalities*. Oxford University Press, 327–349. https://doi.org/10.1093/acprof:oso/9780199832798.003.0014
19. Paul. Ekman, Wallace V. Friesen, and Phoebe Ellsworth. 1972. *Emotion in the human face: guide-lines for research and an integration of findings*. Pergamon Press.
20. Mario Enriquez, Karon MacLean, and Christian Chita. 2006. Haptic phonemes. In *Proceedings of the 8th international conference on Multimodal interfaces - ICMI '06*, 302. https://doi.org/10.1145/1180995.1181053

21. J.B.F. van Erp. 2005. Vibrotactile Spatial Acuity on the Torso: Effects of Location and Timing Parameters. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 80–85. https://doi.org/10.1109/WHC.2005.144
22. Nico H. Frijda. 1986. *The emotions*. Cambridge University Press.
23. Alastair Haigh, David J. Brown, Peter Meijer, and Michael J. Proulx. 2013. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00330
24. Giles Hamilton-Fletcher and Jamie Ward. 2013. Representing colour through hearing and touch in sensory substitution devices. *Multisensory research* 26, 6: 503–32.
25. Eve Hoggan and Stephen Brewster. 2007. Designing audio and tactile crossmodal icons for mobile devices. In *Proceedings of the ninth international conference on Multimodal interfaces - ICMI '07*, 162. https://doi.org/10.1145/1322192.1322222
26. Nicholas. Humphrey. 1992. *A history of the mind*. Chatto & Windus, London.
27. Susan Hurley and Alva Noë. 2003. Neural Plasticity and Consciousness. *Biology & Philosophy* 18, 1: 131–168. https://doi.org/10.1023/A:1023308401356
28. IBM. 2019. SPSS Software. *IBM*.
29. Lynette A. Jones, Jacquelyn Kunkel, and Edgar Torres. 2007. Tactile Vocabulary for Tactile Displays. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, 574–575. https://doi.org/10.1109/WHC.2007.107
30. K.A. Kaczmarek, J.G. Webster, P. Bach-y-Rita, and W.J. Tompkins. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38, 1: 1–16. https://doi.org/10.1109/10.68204
31. Mark L. Knapp, Judith A. Hall, and Terrence G. Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.
32. Li Liu, Shuo Niu, and Mauro Carassai. 2017. The impacts of using different methods to sense emotion in computer-mediated communication. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1214–1219. https://doi.org/10.1109/SMC.2017.8122778
33. J. M. Loomis. 2010. Sensory substitution for orientation and mobility: what progress are we making? In *Foundations of Orientation and Mobility* (3rd ed.), William R. Wiener, Richard L. Welsh and Bruce B. Blasch (eds.). AFB Press, NewYork, 3–44.
34. Jack M. Loomis, Roberta L. Klatzky, and Nicholas A. Giudice. 2013. Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language. In *Multisensory Imagery*. Springer New York, New York, NY, 131–155. https://doi.org/10.1007/978-1-4614-5879-1_8
35. Shachar Maidenbaum and Sami Abboud. 2014. Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews* 41: 3–15. https://doi.org/10.1016/J.NEUBIOREV.2013.11.007
36. Troy McDaniel, Shantanu Bala, Jacob Rosenthal, Ramin Tadayon, Arash Tadayon, and Sethuraman Panchanathan. 2014. Affective Haptics for Enhancing Access to Social Interactions for Individuals Who are Blind. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access* , 419–429. https://doi.org/10.1007/978-3-319-07437-5_40
37. Troy McDaniel, Diep Tran, Samjhana Devkota, Kaitlyn DiLorenzo, Bijan Fakhri, and Sethuraman Panchanathan. 2018. Tactile Facial Expressions and Associated Emotions toward Accessible Social Interactions for Individuals Who Are Blind. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface - MAHCI'18*, 25–32. https://doi.org/10.1145/3264856.3264860
38. Peter Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 2: 112–121. https://doi.org/10.1109/10.121642
39. Peter Meijer. 2019. Seeing With Sound. Retrieved from https://www.seeingwithsound.com/
40. Alva. Noë. 2004. *Action in perception*. MIT Press.
41. Scott D. Novich and David M. Eagleman. 2014. A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In *2014 IEEE Haptics Symposium (HAPTICS)*, 1–1. https://doi.org/10.1109/HAPTICS.2014.6775558
42. J. K. O'Regan. 2011. *Why red doesn't sound like a bell : understanding the feel of consciousness*.

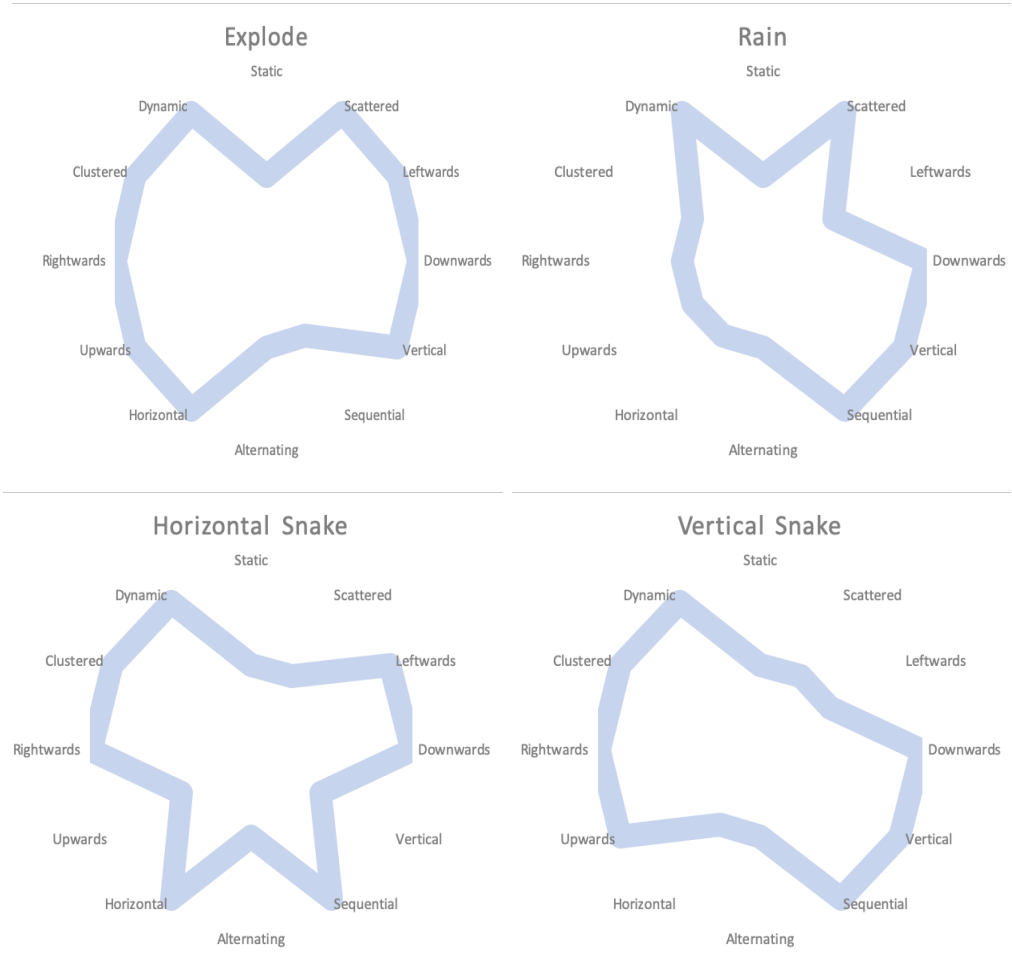
Oxford University Press.
43. Keith Oatley and P. N. Johnson-laird. 1987. Towards a Cognitive Theory of Emotions. *Cognition & Emotion* 1, 1: 29–50. https://doi.org/10.1080/02699938708408362
44. Cesare V. Parise and Charles Spence. 2012. Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research* 220, 3–4: 319–333. https://doi.org/10.1007/s00221-012-3140-6
45. E. Piateski and L. Jones. 2005. Vibrotactile Pattern Recognition on the Arm and Torso. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 90–95. https://doi.org/10.1109/WHC.2005.143
46. Jesse J. Prinz. 2006. Putting the brakes on enactive perception. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 12.
47. Michael J. Proulx, Petra Stoerig, Eva Ludowig, and Inna Knoll. 2008. Seeing 'Where' through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution. *PLoS ONE* 3, 3: e1840. https://doi.org/10.1371/journal.pone.0001840
48. Qualtrics. 2019. Qualtrics. *Qualtrics*.
49. Shafiq Ur Rehman, Li Liu, and Haibo Li. 2007. Facial expression appearance for tactile perception of emotions. 239–242. https://doi.org/https://doi.org/10.1109/MMSP.2007.4412862
50. D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott. 2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences* 107, 6: 2408–2412. https://doi.org/10.1073/pnas.0908239106
51. Michelle N. Shiota, Belinda Campos, Dacher Keltner, and Matthew J. Hertenstein. 2004. *The Regulation of Emotion*. Psychology Press, NewYork. https://doi.org/10.4324/9781410610898
52. Charles Spence. 2011. Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics* 73, 4: 971–995. https://doi.org/10.3758/s13414-010-0073-7
53. Charles Spence and Cesare V. Parise. 2012. The Cognitive Neuroscience of Crossmodal Correspondences. *i-Perception* 3, 7: 410–412. https://doi.org/10.1068/i0540ic
54. Sharmila Sreetharan and Michael Schutz. 2019. Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope. *Multimodal Technologies and Interaction* 3, 1: 4. https://doi.org/10.3390/mti3010004
55. Terrence R. Stanford and Barry E. Stein. 2007. Superadditivity in multisensory integration: putting the computation in context. *NeuroReport* 18, 8: 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315
56. B E Stein and M T Wallace. 1996. Comparisons of cross-modality integration in midbrain and cortex. *Progress in brain research* 112: 289–99. https://doi.org/10.1016/s0079-6123(08)63336-1
57. Barry E. Stein. 1998. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research* 123, 1–2: 124–135. https://doi.org/10.1007/s002210050553
58. Ella Striem-Amit and Amir Amedi. 2014. Visual Cortex Extrastriate Body-Selective Area Activation in Congenitally Blind People "Seeing" by Using Sounds. *Current Biology* 24, 6: 687–692. https://doi.org/10.1016/J.CUB.2014.02.010
59. Ella Striem-Amit, Laurent Cohen, Stanislas Dehaene, and Amir Amedi. 2012. Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron* 76, 3: 640–652. https://doi.org/10.1016/J.NEURON.2012.08.026
60. Akihiro Tanaka, Ai Koizumi, Hisato Imai, Saori Hiramatsu, Eriko Hiramoto, and Beatrice de Gelder. 2010. I Feel Your Voice. Cultural differences in the multisensory perception of emotion. *Psychological Science* 21, 9: 1259–1262. https://doi.org/10.1177/0956797610380698
61. Dzmitry Tsetserukou and Alena Neviarouskaya. 2010. World's first wearable humanoid robot that augments our emotions. In *Proceedings of the 1st Augmented Human International Conference on - AH '10*, 1–10. https://doi.org/10.1145/1785455.1785463
62. Dzmitry Tsetserukou, Alena Neviarouskaya, Helmut Prendinger, Naoki Kawakami, Mitsuru Ishizuka, and Susumu Tachi. 2009. iFeel_IM! Emotion Enhancing Garment for Communication in Affect Sensitive Instant Messenger. In *Human Interface and the Management of Information. Designing Information Environments*, 628–637. https://doi.org/10.1007/978-3-642-02556-3_71

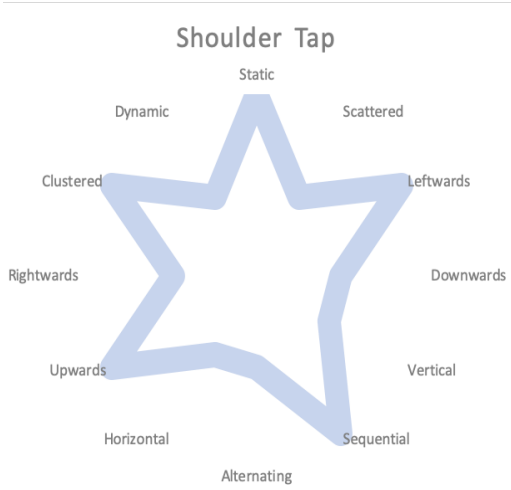
63. Junichi Tsurukawa, Mohammed Al-Sada, and Tatsuo Nakajima. 2015. Filtering visual information for reducing visual cognitive load. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp '15*, 33–36. https://doi.org/10.1145/2800835.2800852
64. Ramiro Velázquez and Omar Bazán. 2010. Preliminary evaluation of podotactile feedback in sighted and blind users. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2103–2106. https://doi.org/10.1109/IEMBS.2010.5626205
65. Ramiro Velazquez and Edwige Pissaloux. 2014. On human performance in tactile language learning and tactile memory. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 96–101. https://doi.org/10.1109/ROMAN.2014.6926236
66. Ramiro Velázquez and Edwige Pissaloux. 2018. Constructing Tactile Languages for Situational Awareness Assistance of Visually Impaired People. In *Mobility of Visually Impaired People*. Springer International Publishing, 597–616. https://doi.org/10.1007/978-3-319-54446-5_19
67. Taekbeom Yoo, Yongjae Yoo, and Seungmoon Choi. 2014. An Explorative Study on Crossmodal Congruence Between Visual and Tactile Icons Based on Emotional Responses. In *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, 96–103. https://doi.org/10.1145/2663204.2663231

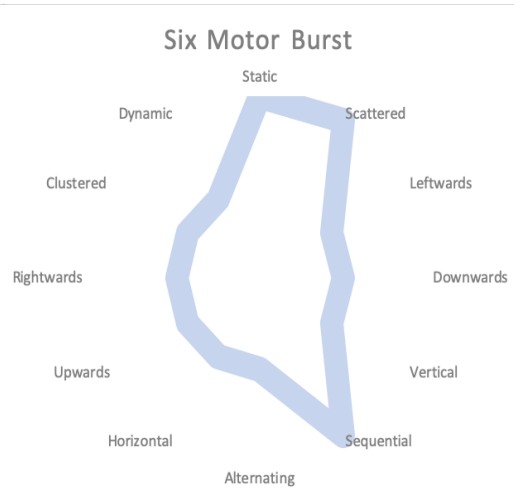
2.8 Appendices

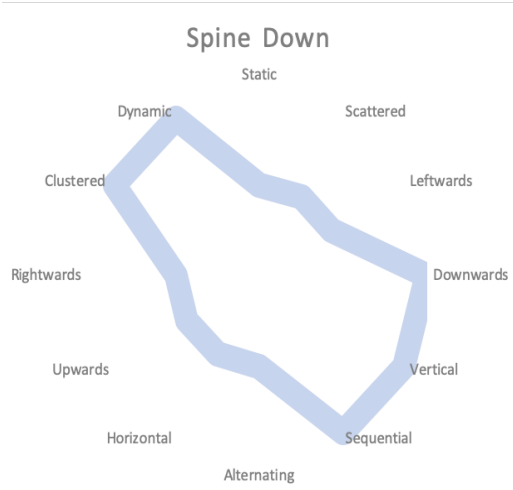
Appendix 2.A: Polar Maps of Stimulation Patterns with respect to Six Spatial Factors

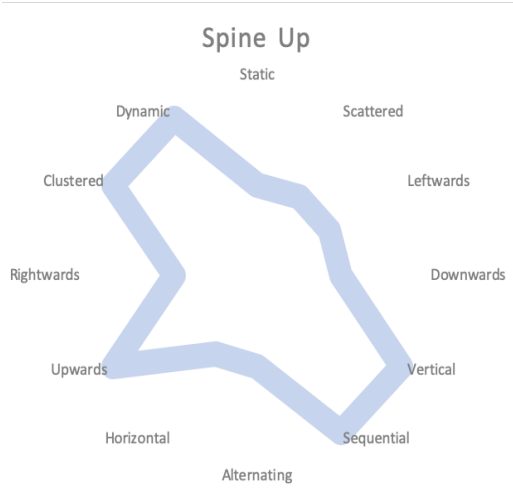
The following represents each stimulation pattern with respect to six spatial factors (static/dynamic, clustered/scattered, rightwards/leftwards, upwards/downwards, horizontal/vertical, alternating/sequential). Static/dynamic refers to whether stimulation frames are different from each other. Clustered/scattered refers to how dispersed the stimulation points are. Rightwards/leftwards, upwards/downwards, and horizontal/vertical refer to how the stimulation frames move across axes. Alternating/sequential refers to whether the stimulation frames alternate between each other.

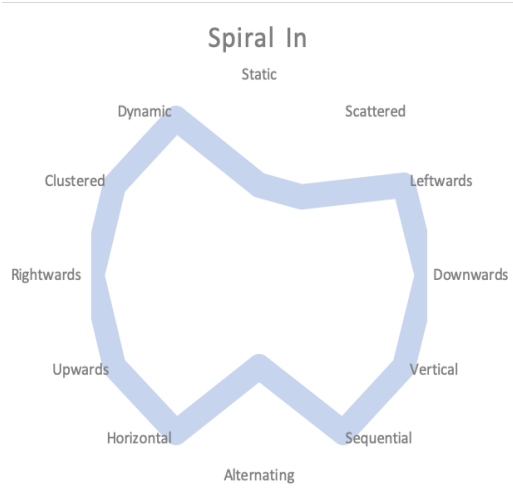


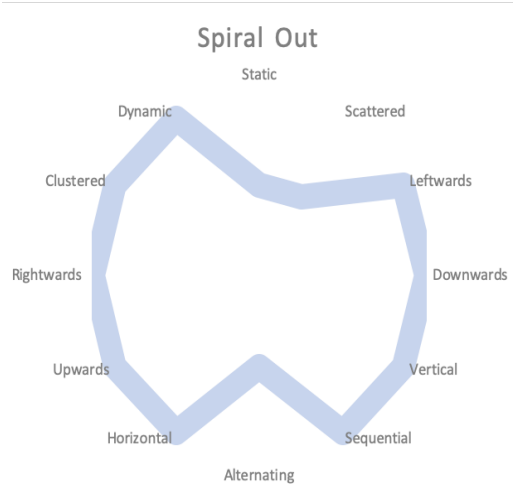

Shoulder Tap
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

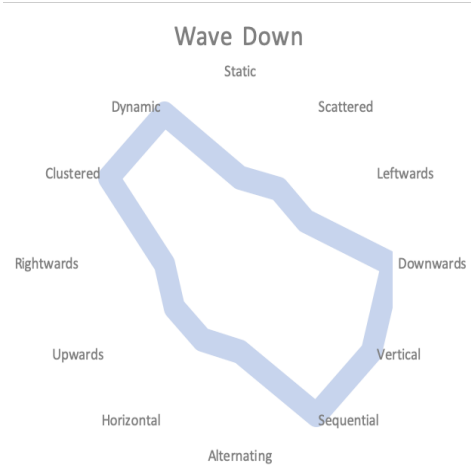

Six Motor Burst
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Spine Down
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

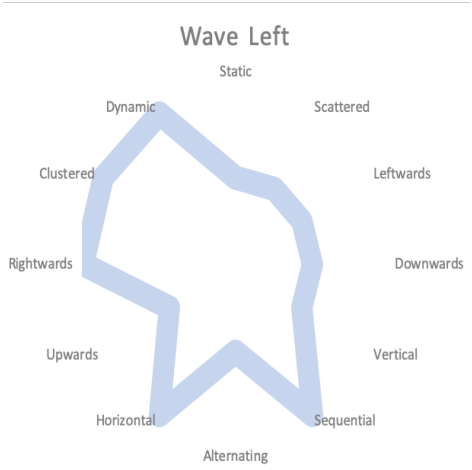

Spine Up
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

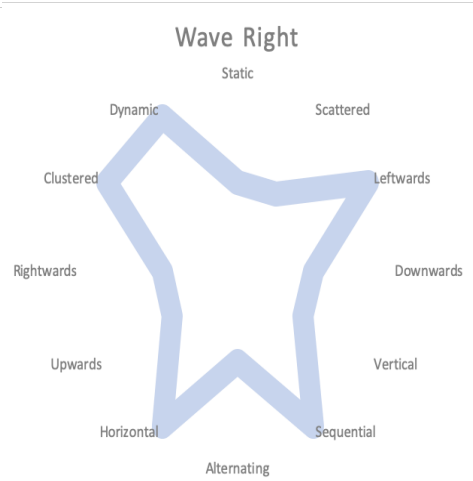

Spiral In
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

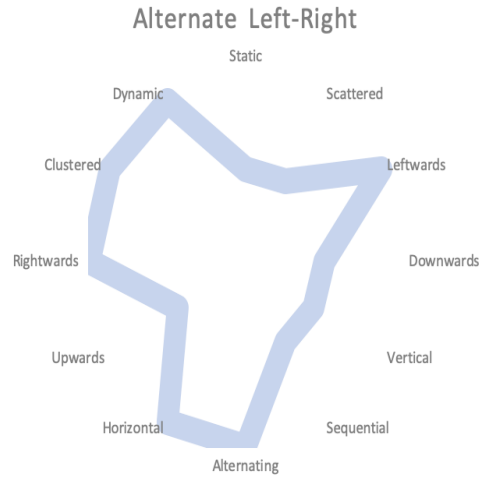

Spiral Out
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

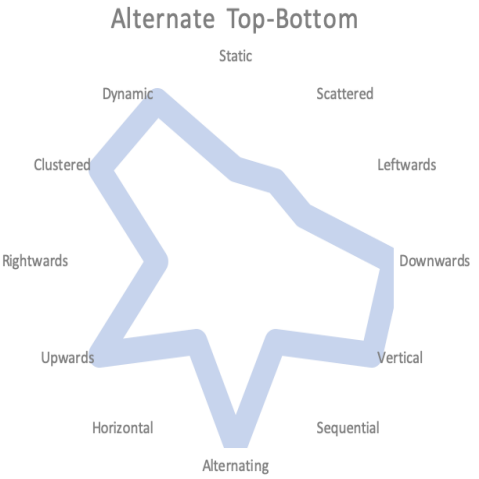

Wave Down
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Wave Up
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Wave Left
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Wave Right
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Alternate Left-Right
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating


Alternate Top-Bottom
Static
Dynamic
Scattered
Clustered
Leftwards
Rightwards
Downwards
Upwards
Vertical
Horizontal
Sequential
Alternating

The contingency table below represents the response frequencies of each stimulation pattern, in percentages, with respect to the emotional response and display mode.

Emotional Response Feedback	Happiness			Anger			Sadness			Surprise			Fear			Disgust			Neutral/Relaxed		
(T: Tactile, S: Sonification, V: Visual):	T	S	V	T	S	V	T	S	V	T	S	V	T	S	V	T	S	V	T	S	V
Alternate Left-Right	3.1	2.5	2.7	12	11.5	15.3	3.4	2.1	0	3.1	1.7	8.4	16.4	13.9	10.7	1.5	8.4	11.5	2.1	3.2	1.1
Alternate Top-Bottom	3.1	3.8	4	3.8	5	15.2	3.4	2.1	0	15.3	6.8	10.8	7.1	14.9	7.8	5.3	4.7	3.8	3.2	2.4	2.2
Explode	4.6	15.3	9.3	3.8	3.3	2.2	4.5	2.1	4.2	7.6	9.6	16.9	4.3	0.6	1.6	6.1	4.7	7.7	4.2	5.6	0
Horizontal Snake	11.3	8.8	8	2.3	2.5	4.3	10.8	4.1	4.2	5.3	5.1	3.6	2.9	7.5	6.3	10.9	4.7	15.7	5.3	9.6	6.7
Rain	2.1	15.7	8	6.8	4.1	0	7.6	3.1	10.2	6.1	6.8	13.5	8.6	2.9	1.6	4.5	5.6	0	6.3	3.2	4.4
Shoulder Tap	3.1	0.6	2.7	5.3	7.4	10.9	3.4	2.1	6.3	6.9	7.3	10.8	5.7	9.8	9.4	8.3	4.7	3.8	4.2	10.4	1.1
Six Motor Burst	1.5	11.9	6.7	10.4	4.1	10.9	2.2	0	2.1	7.6	12	8.4	7.9	3.4	9.4	1.5	4.7	11.5	8.4	3.2	0
Spine Down	11.9	6.9	2.7	1.5	0	0	6.7	25.8	22.9	7.6	2.8	0	5	1.1	4.7	9.8	7.5	3.8	6.3	7.2	11.1
Spine Up	13.9	6.9	8	2.3	5	2.2	10.8	3.1	4.2	6.1	9.5	3.6	5.7	3.4	6.3	6.8	5.6	3.8	4.2	8.8	11.1
Spiral In	7.7	5	4	10.5	5.8	2.2	4.5	5.2	8.3	3.8	7.9	9.6	5	1.1	4.7	9.1	10.2	7.7	11.5	10.4	6.7
Spiral Out	8.8	5.7	8	4.5	3.3	6.5	6.7	11	0	6.1	3.4	4.8	5.7	4.6	3.1	9.8	11.2	3.8	5.3	8	12.2
Vertical Snake	12.4	6.3	9.3	1.5	0	4.3	7.9	5.2	4.2	6.1	8.5	6	5	4	3.1	9.8	5.6	11.5	4.2	13.6	6.7
Wave Down	4.1	3.1	2.7	4.5	2.5	4.3	7.9	25.8	18.8	4.6	6.2	0	2.9	4.6	6.3	9.8	5.6	0	8.4	1.6	11.1
Wave Left	3.6	0	5.3	12.8	17.4	6.5	6.7	2.1	2.1	3.1	2.3	1.2	6.4	9.8	12.5	3	7.5	7.7	5.3	6.4	8.9
Wave Right	3.6	0.6	5.3	12	20.7	8.7	5.6	3.1	8.3	3.8	1.1	1.2	7.1	11.5	7.8	0.8	6.5	7.7	11.6	1.6	7.8
Wave Up	5.2	6.9	13.3	6	7.4	6.5	7.9	3.1	4.2	6.9	9	1.2	4.3	6.9	4.7	3	2.8	0	9.5	4.8	8.9

Reflections III

> "As long as one holds fast to a classical conception of perception in terms of the acquisition of information, one will be stuck with the principle that it is always better to have access to more information. In this framework, persons with sensory handicaps will inevitably be considered as defective…it is the classical perception which carries the germ of exclusion since it considers that the problem of handicapped persons lies in a quantitative difference. By contrast, true respect for the world of handicapped persons lies with better knowledge and understanding of the qualitative difference of possible perceptual modes."[1]

The previous study demonstrated that cross-modal tactile or sonification display modes, and their visual equivalents could convey basic emotions over digital communication systems. It also suggested future research directions to conceptualise cross-modal associations between stimulation patterns and emotions for inclusive display development. One of the research questions the study addressed was, which sensory representations formed the basis of *unisensory* perception with cross-modal displays? Gaining an understanding this contributes significantly to the inclusive applications of sensory substitution techniques because they can, in principle, grant access to the same sensory information via different sensory channels.

For cross-modal displays and sensory substitution techniques to be used in an inclusive context, their perception should be understood deeply. In this way, cross-modal displays can serve in inclusive display development as long as the sensory forms they transform are studied in more detail. The study presented in the previous chapter found that cross-modal display modes did not always lead to the same emotional responses (e.g. positive emotions), despite carrying equivalent forms of stimulation

[1] Lenay C., Gapenne O., Hanneton S., Marque C., Genouëlle C., Sensory Substitution: Limits and Perspectives, 2003.

patterns. This was explained by (i) how the cues for positive emotions might be culture-specific, and (ii) perception with cross-modal displays utilises multiple senses. The latter also indicates that perception with cross-modal displays might fall between the substituting and substituted senses. These raised the question of to what extent cross-modal feedback of different sensory origins could be perceived to carry the same form. It was also preliminarily suggested that the perception with cross-modal display modes might be biased towards the way in which the stimulation patterns move across different dimensions.

The upcoming chapter will address some of these questions with the evaluation of a cross-modal display prototype with unisensory and multisensory modes. Since conveying emotions might carry a certain degree of subjective bias, an object recognition task, which is commonly studied with various sensory substitution techniques, will be further utilised. The next study also offers a methodology to study cross-modal displays in a multisensory context. By doing so, it is aimed to investigate whether cross-modal displays with multisensory modes could enrich our interaction with the digital world. The next study further expands the investigation of which sensory representations form the basis of *multisensory perception* with cross-modal displays. Given that different cross-modal display modes might carry dimension-specific perceived resolution, the next study will further discuss its outcome. Overall, the next study aims to address whether multisensory combination or integration occurs while using cross-modal displays with multisensory modes.

Shortly after the analysis of the previous study was completed, our grant application to acquire BrainPort was approved by the University of Bath Alumni Fund. Here we have the chance to deeply thank Bath Alumni to support our research in HCI and cognitive sciences. Having BrainPort, a tactile-to-visual sensory substitution device, in the Cross-Modal Cognition Lab was quite exciting. Consequently, the following two chapters will utilise BrainPort's intra-oral interface as an electrotactile display mode in making cross-modal display prototypes with unisensory and multisensory modes.

CHAPTER III

Two Is Better Than One: Multisensory Combination of Auditory and Tactile Cross-Modal Displays

Declaration

<table>
<tr><td colspan="4">This declaration concerns the article entitled:</td></tr>
<tr><td colspan="4">Two Is Better Than One: Multisensory Combination of Auditory and Tactile Cross-Modal Displays</td></tr>
<tr><td colspan="4">Publication status (tick one)
Draft manuscript [x] Submitted [ ] In review [ ] Accepted [ ] Published [ ]</td></tr>
<tr><td>Publication details (reference)</td><td colspan="3">N/A</td></tr>
<tr><td colspan="4">Copyright status (tick the appropriate statement)
I hold the copyright for this material [x] Copyright is retained by the publisher, but I have been given permission to replicate the material here [ ]</td></tr>
<tr><td>Candidate's contribution to the paper (provide details, and also indicate as a percentage)</td><td colspan="3">The candidate predominantly executed the
Formulation of ideas: 90%
Design of methodology: 90%
Experimental work: 75%
Presentation of data in journal format: 90%
For details, please see acknowledgements on the next page</td></tr>
<tr><td>Statement from Candidate</td><td colspan="3">This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.</td></tr>
<tr><td>Signed</td><td></td><td>Date</td><td></td></tr>
</table>

Acknowledgements

I thank Izzy Fitton, Grainne Bannigan and Michael Colman for their help with the data collection. I am grateful to Michael Proulx and Eamonn O'Neill for their discussion of the research. I am also grateful to Vanessa Lloyd-Esenkaya for patiently proof-reading the manuscript. I also thank all the participants who took part in the experiment.

3.0 Abstract

Even though the human brain has evolved in a multisensory environment, the visual pixels have been dominating the access to the cyberworld. Semantic inconsistencies surrounding multisensory processing and interdisciplinary research have been blamed for this as an obstacle in developing and assessing multisensory systems. To bridge the gap in research, the current research reviews a framework of multisensory processing based on cognitive principles. This framework is then operationalised in prototyping a cross-modal display with unisensory (auditory or tactile) and multisensory (audio-tactile) display modes to further investigate the types of multisensory processing and how they occur while using multisensory display modes. In a study examining object recognition (N = 48), the performance of the unisensory and multisensory display modes were investigated. The findings revealed that the multisensory display mode resulted in the highest performance as a result of multisensory combination, a type of multisensory processing. Overall, by applying a multisensory framework, the current research successfully demonstrates how a mixed study design enables the improvement of cross-modal displays and identifies HCI research areas that can progress the development of inclusive technologies.

3.1 Introduction

The human brain has evolved in a way that is crucial for a coherent multisensory experience [18,94]. We have a unified system of senses to perceive and act upon this environment [29,34,95]. With the advent of digital technologies, however, we now "live between two realms: our physical environment and cyberspace" [46]. While we are inherently adept at interacting seamlessly with the rich multisensory information channels from the physical world, our interactions with cyberspace are artificially dependent on graphical user interfaces (GUIs). Consequently, relying on GUIs drifted us away from interacting with cyberspace with our five senses as we do seamlessly with the physical world. This had led to two main challenges in human-computer interaction (HCI) research: (i) creating rich multisensory experiences in new technologies such as virtual reality (VR) and augmented reality (AR), and (ii) building inclusive technologies. One approach to address these challenges is first to rigorously explore non-visual interactions in HCI and then unify them in displays with multisensory modes. Indeed, as to the first part of this approach, there are many studies investigating auditory [50,82] and tactile (see Haptipedia [85]) interface designs, and chemical senses (i.e. smell and taste) as a means of interacting with cyberspace [87]. For the purposes of this paper, we will focus on auditory and tactile sensations in the context of examining multisensory systems.

Despite the extensive and careful investigation of the design of unisensory interfaces (e.g. visual, tactile or auditory display modes), relatively fewer studies have so far looked at multisensory systems in HCI [88]. This might be because a number of studies evidenced that unisensory interfaces could outperform multisensory ones due to the increased cognitive load of the latter [55,77,86]. A clear scientific agreement on this matter, however, remains elusive [58]. Moreover, suggesting that multisensory systems are inferior to their unisensory alternatives is especially surprising given that studies grounded in psychology and neuroscience repeatedly illustrate that multisensory phenomena enrich our interactions with the physical world [18]. Instead, a more plausible reason why multisensory systems are not as eminent in HCI might be because of how they are designed and studied. Semantic and methodological

inconsistencies among different disciplines targeting multisensory phenomena have indeed been blamed for this [93]. That is, progress in studying and utilising multisensory phenomena depends on a common lexicon and robust methodologies to successfully transfer evidence from behavioural and neural studies to other applied domains of research.

If multisensory phenomena enrich our interactions with the physical environment, then developing displays with multisensory modes should improve how we interact with cyberspace. Applying multisensory theory to practice, however, is challenging because HCI research has not yet found effective ways to successfully mirror multisensory cognitive functions in interaction design [74,75]. Our paper therefore aims to address some of these challenges in developing displays with multisensory modes. First, we review a framework of established principles of multisensory processing to be used as design patterns for display development. Next, we implement this framework in our own experiment, which pairs auditory and tactile cross-modal display modes, to demonstrate how the framework can be successfully utilised in HCI research. Overall, we investigate whether multisensory combination or integration occurs while using cross-modal displays with multisensory modes. Our findings evidence that it is the complementary use of sensory cues (i.e. multisensory combination), and not their redundancy (i.e. multisensory integration), that enhances user performance via novel cross-modal displays with a multisensory mode. Overall, we contribute new insight in developing and examining multisensory display modes, particularly with cross-modal displays, and thus improve their successful use and inclusion.

3.2 Background

The theoretical framework provided here is independent of sensory origin and can therefore be generalised across the development of various multisensory display modes (e.g. audio-tactile, audio-visual, tactile-visual). The current research based on this framework introduces a novel audio-tactile cross-modal display prototype, which

combines auditory-to-visual and tactile-to-visual sensory substitution techniques. That is, the display prototype could take in visual input and then translate the visual feed into something that is heard and/or touched. In making the prototype, two sensory substitution devices, namely the auditory vOICe [67] and tactile BrainPort [105], were utilised. The vOICe and BrainPort are successful exemplars of auditory and tactile cross-modal displays, studied rigorously in multiple domains of research and used for various forms of rehabilitation such as vestibular [21] or visual impairments [37,63].

3.2.1 A Framework for Multisensory Processing

Building a robust theoretical framework is important and requires a consistent nomenclature. Therefore, the key concepts and principles of multisensory processing are defined here. Stein et al. [93] documented a guideline for such a common nomenclature for multisensory phenomena. Multisensory describes a neural or behavioural process associated with multiple senses. To deal with the semantic inconsistencies relating to multisensory phenomena, they suggested using multisensory processing as a generic term. Accordingly, multisensory processing refers to any multisensory phenomenon such as multisensory integration or multisensory combination. Unisensory refers to any neural or behavioural process associated with a single sense. Modality-specific is a stimulus property confined to such unisensory processes. C*ross-modal* refers to a complex of two or more modality-specific stimuli. Finally, cross-modal matching is a cognitive process in which cross-modal stimuli are compared to estimate a multisensory equivalence of the sensory source.

According to these definitions, some concepts (i.e. modality-specific and cross-modal) are attributed to the properties of the sensory source while others (i.e. unisensory and multisensory*)* to our neural or behavioural responses. In line with these, a multisensory technology (e.g. a display mode) should evoke a multisensory response from the users by delivering cross-modal cues. Therefore, understanding multisensory processing in a framework is key to developing displays with a multisensory mode.

Based on these definitions, there are a number of principles for multisensory processing to occur [54]. These concepts and principles will be later attributed to the empirical methodology for studying multisensory display modes in the current research (see Experimental Investigation).

3.2.1.1 Principles of Spatial and Temporal Coincidence

The spatial coincidence principle highlights that the cross-modal information should be collected from spatially aligned sensory sources to enhance multisensory response quality [92]. The temporal coincidence principle similarly suggests that cross-modal information that is received from close temporal proximity would improve multisensory response quality [91]. It is argued that cross-modal information that is not spatially or temporally aligned might be perceived as if they come from separate sensory sources. This causes depression in the multisensory response and instead leads to separate unisensory responses [91,92].

Principles of spatial and temporal coincidence are manipulated in a number of HCI studies to investigate the margins between congruent and incongruent cross-modal cues with respect to user experience and performance. For example, incongruent audio-visual stimuli could be implemented to prevent issues related to distance compression in VR [31]. Congruent visual and audio/tactile stimuli could also be used to increase the perceived quality of buttons on touch-screen devices [43]. In contrast, incongruent audio-visual stimuli could negatively impact target localisation [20]. Despite their contribution to HCI research, these studies examine the principles of spatial and temporal coincidence as sensory (in)congruence. Such conceptualisations point out some of the semantic inconsistencies that should be noted for transparency between multisensory research and HCI.

3.2.1.2 Principle of Inverse Effectiveness

The inverse effectiveness principle refers to how reliable cross-modal cues are in cross-modal matching. That is, if one modality-specific cue elicits a stronger

behavioural response than the other modality-specific cue when presented together from the same sensory source, this would weaken the multisensory processing of the sensory source [78,89–91]. In other words, for multisensory processing to occur, cross-modal cues should be equally reliable. The multisensory calibration, thus the reliability, of our intact senses usually emerges at different critical periods during development [12]. Acquired senses (e.g. via sensory augmentation devices) could be similarly calibrated with intact senses. The reliability of acquired senses could be improved for multisensory processing through learning and experience [80]. This also implies that the users of augmentation displays would rely heavily on their intact senses until the acquired sense becomes equally reliable over time. If more than one novel display mode is used simultaneously, however, their reliabilities are expected to be of equal value to each other, given that they carry the same informational capacity.

3.2.1.3 Multisensory Integration

Among the semantic inconsistencies, multisensory integration, and multimodal as its synonym, is identified to be the most problematic due to its wide use in different bodies of research [93]. This creates a challenge, especially in researching multisensory display modes, if displays are designed without addressing the relevant literature. It inevitably causes an invisible gap between theoretical and applied research. For example, multisensory display modes intended for multisensory integration could as well benefit from other multisensory processes such as multisensory combination. That is, if multisensory systems are evidenced to be inferior to their unisensory alternatives, this might be because they are not designed to take full advantage of multisensory processing. To avoid these pitfalls in the current methodology, multisensory integration will be reviewed in terms of the principles of spatial and temporal coincidence and inverse effectiveness.

Multisensory integration refers to a neural process in which cross-modal cues are integrated to produce a multisensory response that is significantly different from the unisensory response [93]. For example, the size of an odd pear can be estimated by only touching it, only seeing it or inspecting it using both of these senses. Assuming

these senses are equally reliable, if the pear looks to be the size of 6 units and feels like it is the size of 8 units, when it is seen and touched simultaneously, multisensory size will be of 7 units. It is therefore suggested that once both senses are used for inspection, the multisensory estimate will be somewhere between the individual unisensory estimates [83]. The unisensory estimates are thought to be aggregated into a weighted multisensory estimate, where the weight of each unisensory estimate is in proportion to the reliability of the modality-specific cue (i.e. principle of inverse effectiveness) [84]. It is therefore argued that multisensory integration occurs in a statistically optimal fashion and this statistical optimisation follows the rules of maximum likelihood estimation (MLE) [28].

Multisensory integration can be viewed as the neural process of combining redundant sensory cues in an optimal fashion [84]. In this definition, redundancy requires the cross-modal cues to represent the same properties of the sensory source that are in line with the principles of spatial and temporal coincidence. When a discrepancy occurs between the cross-modal cues, MLE would predict that more weight would be attributed to the more reliable cross-modal cue. For example, if the hypothetical pear discussed appears to be the size of 4 units visually, due to heavy fog, and still feels like it is the size of 8 units, the tactile estimate would be more reliable according to the principle of inverse effectiveness. In theory then, the size of the pear will be perceived to be between 7 and 8 units. If the discrepancy between the cues is too high, multisensory integration would not occur at all. This could happen when there is reduced visual field in heavy fog. In this instance, the pear will be perceived to be 8 units, despite sensory feedback being received from both senses. Moreover, the shift in statistical weights, hence their reliability, is adjusted as an automatic process during multisensory integration [100].

3.2.1.4 Multisensory Combination

Multisensory combination is essentially different to multisensory integration as it does not require the principles of spatial and temporal coincidence and inverse effectiveness to the same extent [29]. It is the behavioural and cognitive process of

using complementary cross-modal cues that maximises the overall multisensory estimate (instead of aggregating the unisensory weights of the redundant information) [16]. As cues are not redundant, MLE optimal integration approach cannot predict the outcome of multisensory combination [84]. For example, the shape of an object can be consciously estimated better by visually inspecting it from front and also touching it from behind where it would be invisible to the eyes [70]. In this instance, the cross-modal cues are not redundant. Instead, they complement each other to create a more robust representation of the sensory source. When multisensory integration is inhibited (e.g. due to high discrepancy between cues), multisensory combination is able to combine available cues to estimate the overall size. In this way, multisensory combination might be an additive process.

Our research into multisensory HCI finds that while most studies aim for multisensory integration [93], complementary cross-modal cues are indirectly used to decrease the cognitive load of the users [42]. Delivering the theoretical literature in multisensory combination to a wider audience will bring novel perspectives into developing multisensory display modes. For example, displays inspired by multisensory combination could reach wider adoption by various user profiles and be applied in diverse use cases. This is harder to achieve by displays that aim for multisensory integration as the reliability of the cues would require longer calibration periods. Displays on the basis of multisensory combination could also be used in a wider context by users as they are not bound to the principles of spatial and temporal coincidence.

3.2.2 Sensory Substitution

Sensory substitution is a cognitive process in which one modality-specific cue and its unisensory response (e.g. seeing) can be represented by another (e.g. via touch or hearing) [9,67]. Research in this domain has revealed more insight about the multisensory processing and cross-modal matching, and how this insight can be incorporated in HCI [30]. Sensory substitution devices (SSDs) have been evidenced to be a successful assistive technology with applications in visual [37,60,63,81], hearing

[17,23,53,72] and vestibular [21] domains. These same devices have also been considered as sensory augmentation devices, whereby a person with no sensory impairments could augment their abilities with additional inputs such as thermal imaging [69]. The evidence for their success further suggests that sensory substitution techniques could be studied outside rehabilitation purposes towards building inclusive technologies.

By deploying sensory substitution techniques, for example, the challenges of GUIs outlined earlier can be addressed [32,41]. That is, GUI-dependent interactions can be translated into auditory [101] or tactile [19,47] forms so that they can be presented in the most appropriate multisensory context to enhance user experience. Consequently, in recent years, there has been an increasing interest, both commercial and scientific, in facilitating sensory substitution techniques in designing inclusive displays with multisensory modes. This also highlights our scientific motivation for studying sensory substitution techniques in our empirical study.

3.2.2.1 Sensory Substitution Devices as Cross-Modal Displays

Sensory substitution devices are essentially cross-modal displays that are built, in principle, with how complementary cross-modal cues correspond to each other [52]. There are two mainstream types of SSDs with a unisensory mode available. They convert visual representations into either auditory or tactile cues. The vOICe [67] and BrainPort [105] are successful exemplars of such auditory and tactile cross-modal displays respectively, which are also commercially available. They both consist of three main parts: a camera to capture visual cues, a processor unit to convert the live camera input into auditory or tactile feedback, and an output device (stereo headphones for The vOICe and an intra-oral interface for BrainPort).

The vOICe

The vOICe algorithm converts visual pixels from the live camera input by encoding the position and brightness of pixels as a function of auditory pitch and loudness. Pixels

higher on the y-axis are converted into higher-pitched signals, and brighter pixels are converted into louder signals. The vOICe scans an image from left-to-right and consequently produces a temporal signal to map the pixels on the x-axis via stereo headphones [67]. The final signal carries 11,264 auditory pixels [37]. These soundscapes are meaningless when heard, yet even novice users could reconstruct them in a visuospatial context [38].

A study with a group of congenitally blind and sighted adults showed that participants learned to recognise a range of images including letters, textures, faces, houses, objects, body shapes and geometric shapes [98]. Overall, this study suggested that The vOICe users can learn to recognise even complex visual stimuli, such as faces, after being trained for an average of 73 hours. Contrary to the extensive training applied in the previous study, a number of other studies have shown that novice sighted users can also successfully interpret the soundscapes from The vOICe after shorter training periods [7,96,97]. For example, one study found that blindfolded participants were able to identify objects with 88% success rate after 3 hours of training [7]. Another study showed that novice vOICe users were able to complete an object recognition task with 58.8% success rate after 30 minutes of training [10]. These studies therefore suggest that while The vOICe is able to help users recognise shapes, its performance is correlated with training duration.

BrainPort

BrainPort spatially aligns the visual pixels from the live camera input with the electrodes placed on a 20x20 matrix on the intra-oral interface [105]. This also means that the perceptual resolution of the intra-oral interface is reduced to 400 pixels by default. The brightness of visual pixels is represented with the stimulation intensity of electrodes. The brighter the pixel, the higher the intensity is. An analogous example to the experience of using BrainPort could be when someone draws a shape on your palm and asks you to recognise it without visual help.

In one study, blind users were trained with BrainPort for 10 hours and tested for their object (shapes and letters) recognition performances in every three months over a year [36]. At the end of the year, success rates of participants in shape recognition tasks were at 91.2% whereas word and letter recognition success rates were at 57.9%. In a recent review of BrainPort, it was further reported that training duration (varied between 30 min and 15 hours) and performance in recognition tasks (success rates varied between 15% and 91%) were significantly correlated to each other [99].

3.2.2.2 Multisensory Use of Auditory and Tactile Cross-Modal Displays

The informational capacity of our senses varies greatly in resolution. The eye is estimated to convey information in $4.3x10^6$ bps [49], the ear in 10^4 bps [48] and fingertip in 100 bps [56]. Moreover, The vOICe carries 11,264 auditory pixels [37] and BrainPort is designed to deliver approximately 400 tactile pixels on the tongue [105]. These variations in resolution indicate that the eyes have superior informational capacity to auditory and tactile senses. In the context of sensory substitution techniques, moreover, it is argued that auditory cross-modal displays would deliver higher bandwidth of sensory information [37]. If this is true, auditory cross-modal displays might be expected to yield higher performance than their tactile alternatives. On the other hand, such variations in informational capacity are bounded by the sensory organ or device, hence modality-specific, without entailing the perceived resolution of the unisensory or multisensory response. Considering how multisensory processing can enrich our perceptual judgements [18], it could be expected that multisensory responses would have higher informational resolution. Multisensory processing can lead an additive [90], even multiplicative, increase in perceived information capacity.

To the best of our knowledge, cross-modal displays with multisensory modes have not yet been studied in the context of whether their perceived information capacity might be superior to their unisensory alternatives. While there has been no attempt to explore them for generic purposes (i.e. navigation and recognition), a few studies investigated cross-modal displays with multisensory modes, which were specialised in

navigation to augment the white cane. EyeCane is a hand-held SSD and augments the capabilities of the white cane by delivering point-distance information via the frequency of vibration and/or auditory cues [1,14,64,65]. It was demonstrated that, having trained for less than 5 minutes, both visually impaired and sighted (blindfolded) EyeCane users successfully completed a variety of spatial tasks and outperformed a control group who were only given a white cane. A more recent cross-modal display prototype with a multisensory mode, namely SoV (Sound of Vision), works similarly by encoding depth and direction information into auditory cues and by delivering the direction of the closest object via a haptic belt [40,51]. This research also showed that blind SoV users' ability to perform spatial tasks improved after 8 hours of training. Comparing SoV to white cane, however, it was found that users performed the best with the cane.

There are differences between EyeCane and SoV which may be rooted in the semantic inconsistencies and the lack of a mutual methodology for testing multisensory display modes. Despite conveying auditory and tactile feedback, EyeCane was not referred as a multisensory system. It was rather described as a "minimal sensory substitution device" for its small size and short training requirements [15]. In researching SoV, on the other hand, it was concluded with the "challenges of multisensory integration" should be addressed [40]. In this research, which clearly aimed for multisensory integration, it was not made clear whether SoV was designed with respect to the principles of multisensory processing. As SoV encodes different properties of the sensory source (i.e. modality-specific), it might be the case that SoV does not provide redundant sensory information by design. If this is the case, to enhance user performance, SoV could benefit more from research which investigates multisensory combination rather than research addressing the challenges of multisensory integration. Similarly, the reasons why the white cane outperformed SoV could be explained by the principle of inverse effectiveness. Overall, these issues accentuate how operationalising a theoretical framework could be rewarding both for the multisensory HCI research and also the potential users.

3.3 Experimental Investigation

In the following empirical study, the methodology operationalises the key principles of multisensory processing as a theoretical framework and provides a robust way of studying cross-modal displays with multisensory modes. Firstly, to be consistent with the principles of spatial and temporal coincidence, an auditory-tactile cross-modal display prototype, namely Cross-Modal Box, was developed to spatially and temporally align the sensory cues. By controlling the properties of the sensory source (cross-modal as opposed to modality-specific), behavioural responses were examined as user performance from unisensory (i.e. auditory or tactile) and multisensory (i.e. audio-tactile) display modes in an object recognition task. In making the Cross-Modal Box, The vOICe and BrainPort were utilised as cross-modal displays. By doing so, Cross-Modal Box was designed to convey redundant cues as sensory substitution techniques would enable the delivery of the same information via different sensory representations.

The current literature suggests that redundant information is processed by multisensory neurons whose overlapping receptive fields correspond to different sensory input such as tactile and auditory [91, 92]. While it is evidenced that this results in faster reaction times in multisensory integration [35,39], the exact principles of multisensory combination and its influence on reaction times are yet to be examined in more detail [29]. Response accuracy, on the other hand, has been shown to improve in both multisensory integration and combination [26,33,66,102]. To study whether multisensory combination or integration occurs while using cross-modal displays with multisensory modes, participants' verbal responses and reaction times were recorded as measures of response accuracy and reaction times.

Secondly, in line with the principle of inverse effectiveness, only the participants who were novice to the Cross-Modal Box, The vOICe and BrainPort were recruited. The information capacity which The vOICe and BrainPort were able to deliver was also controlled by equalising their resolutions at 400 pixels. Thus, the cross-modal displays delivered the same redundant information, avoiding any large discrepancy between

the two unisensory estimates that might otherwise be biased towards the more reliable one. By using basic 2D objects (shapes and letters) as stimuli to assess performance in an object recognition task, it was aimed to minimise the cognitive load participants might experience due to interacting with novel devices, and complex objects and tasks. Additionally, as The vOICe and BrainPort were studied extensively in object recognition [76], deploying a similar task formulated a basis of comparison with the multisensory mode of Cross-Modal Box. As multisensory processing is modulated by attention [100], participants were also blindfolded to avoid visual distraction and reliance so that they could solely focus on the auditory and/or tactile cues.

Having users' qualitative strategies along with quantitative measurements revealed significant insights about how novice and expert users experience cross-modal displays such as The vOICe [7,10,38,103] and BrainPort [4,22,27,36]. To date, very few studies of multisensory display modes and cross-modal displays have conducted semi-structured interviews and collected user feedback. This means that how users utilise multisensory information channels is not studied in depth. To address this, participants in the current study were asked to report the strategies they used for unisensory and multisensory display modes.

With a within-subject and mixed study design of quantitative and qualitative methodologies, the following empirical study demonstrates how unisensory and multisensory display modes could be examined in the current framework. It also exemplifies how user feedback could be incorporated into the improvement of multisensory systems. The wider goal of this research is to operationalise multisensory principles in cross-modal display development. In this way, it is aimed to enrich the user experience with tangible interactions and extended reality technologies, such as virtual and augmented reality, thereby improving their inclusive reach. As the first step towards this goal, in the current research, only adults with no visual, auditory or tactile impairments were studied.

3.3.1 Experimental Hypotheses

It was hypothesised that the multisensory mode of Cross-Modal Box would improve overall task accuracy (H1.1) and speed up reaction times (H1.2) in object recognition. Furthermore, it was hypothesised that providing redundant sensory cues and following multisensory principles in the methodology would lead to multisensory integration (H2.1). Multisensory integration would consequently formulate the basis of enhanced performance predicted in H1. Following this, it was also hypothesised that user strategies should reflect the benefits of multisensory processing (H2.2).

3.4 Methods

3.4.1 Participants

In total, 48 participants (24M, 46 right handed) from 18 to 38 years of age (M = 22.0, SD = 4.72) were recruited from the University of Bath, UK. They were novice to Cross-Modal Box, The vOICe and BrainPort. All participants reported normal or corrected vision, normal hearing and tactile sensation (e.g. they did not have any cold, and intra-oral inflammations and cuts). In this way, hygiene was maintained between participants, and it was ensured that participants would be able to use Cross-Modal Box comfortably and efficiently. The experiment lasted for approximately 60 minutes, and participants were reimbursed £5 for their time. The study was approved by the University of Bath ethics committee (reference number 17-204) and all participants provided consent prior to the study onset.

3.4.2 Cross-Modal Box

Cross-Modal Box is an audio-tactile display prototype that enables participants to explore the given stimulus via sonifications, two-dimensional electrotactile cues and their multisensory equivalent. For this prototype, The vOICe and BrainPort were utilised as auditory and tactile cross-modal displays respectively. Unlike The vOICe, which is purely a software and compatible with most camera connected devices,

BrainPort is intact and works only in real time. That is, it consists of its own helmet, camera and processor, and therefore cannot be connected to an external camera or loaded with pre-recorded stimulus. This technically prevented Cross-Modal Box to convey the same stimulus in the multisensory mode from one live camera feed. Even if two cameras (one for each cross-modal display) were to be placed adjacently and used simultaneously, they would misalign the cross-modal cues spatially and temporally. As this misalignment would lead to multisensory response depression [91,92], using two separate cameras was not considered. Consequently, Cross-Modal Box was designed to deliver pre-recorded sonifications and real-time electrotactile cues in unisensory and multisensory modes.

Overall, Cross-Modal Box consisted of an enclosed body (40x40x40cm) to control environmental factors such as lighting, an adjustable scaffolding mechanism to attach BrainPort's camera, and a PC to run the unisensory displays simultaneously (Figure 3.1). Inside Cross-Modal Box was an A5 sized (14.8 x 21 cm) viewing platform where a stimulus can be placed 23 cm away from the camera. The tactile display mode (i.e. the intra-oral interface and its cord) were disinfected between participants. They were initially wiped with a cotton ball soaked with 70% isopropyl alcohol and then left in the alcohol solution for 30 minutes. Afterwards, they were cleaned with water and air-dried. A medical bin was used for disposal.

3.4.3 Stimuli

In total, two sets of stimuli were created. The training set consisted of four lines and five circles in horizontal, vertical, ascending and descending orientations (Appendix 3.A). The experimental set consisted of eight shapes (square, rectangle, triangle, right hand triangle, diamond, hexagon, circle, star) and eight letters (J, E, K, Q, R, S, C, Z) (Appendix 3.B). The stimuli were created using Keynote [2] and had the same format (i.e. centred, white stimulus on a black background, Helvetica Neue font and 800pt). These stimuli were printed on an A5 paper to be placed on the viewing platform inside Cross-Modal Box.

Figure 3.1 shows a user with Cross-Modal Box in the multisensory mode. The user has the intra-oral interface on her tongue for electrotactile cues and bone conduction headphones for sonifications.

In line with the inverse effectiveness principle, it is important that cross-modal cues are equally reliable so that their estimates are equally weighted for optimal integration. Therefore, adding noise to one or both of the cues is a common manipulation in researching multisensory processing, particularly multisensory integration and sensory (in)congruence [31]. This is done so that the reliability of one modality-specific cue is reduced to that of the other cue to prevent sensory dominance. Given the information capacity of The vOICe is orders of magnitude higher than BrainPort's, the auditory display mode could dominate the tactile mode when they are used together [37]. Considering that the sensory dominance of one display mode over the other would prevent the benefits of the multisensory mode, the information capacity of the auditory display mode was equalised with the tactile mode at 400 pixels (Figure 3.2). In order to achieve this, BrainPort's HTML based interface was used to save images of the stimuli as they would appear on the intra-oral interface. By sonifying these images, which were already in 400 pixels, with The vOICe, it was aimed to equalise the reliability of both displays. Nonetheless, this should not negatively affect the performance with The vOICe. It was demonstared that

participants novice to The vOICe were able to successfully recognise sonified objects even when the resolution of the source images were reduced to 64 pixels [13].

The design of Cross-Modal Box ensured that the pre-recorded sonifications and tactile cues were identical in their visual origins. Moreover, the auditory and tactile display modes were run simultaneously at exact configurations across participants. That is, sonifications were created at default settings of The vOICe (1s scan rate, normal contrast, foveal view off) and delivered via Docooler stereo bone conduction headphones. Using bone conduction headphones helped the participants hear the instructions from the experimenter and report their verbal responses. The tactile stimuli were delivered via BrainPort's intra-oral interface at the following settings: 37° zoom, low light, high contrast and 18° tilt. The intensity of electrotactile stimuli was initiated with 50% and adjusted for each participant's comfort (average intensity across participants was 60%, SD = 12.0%).

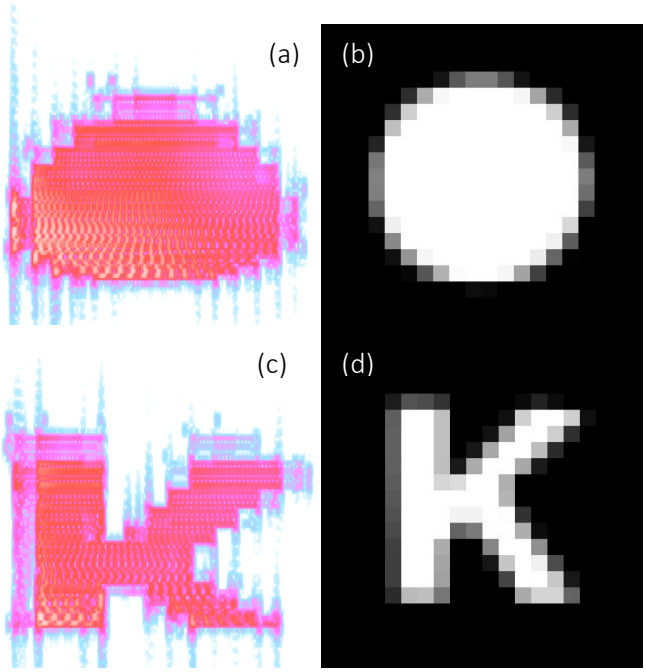


Figure 3.2 exemplifies a shape (i.e. circle) and a letter (i.e K) stimuli. Images on the right (b & d) display the tactile representations conveyed via the intra-oral interface. These images were saved and then sonified with The vOICe to create the auditory stimuli. The spectrograms on the left (a & c) represent the spectrum of frequencies of these sonifications.

3.4.4 Experimental Conditions

In total, there were two unisensory and one multisensory experimental conditions. Each condition consisted of the same set of trials (16 objects) and the trials were presented in a random order. The conditions were further counterbalanced in six variations, and an equal number of male and female participants were assigned to each.

3.4.5 Procedure

Prior to the onset of the training, participants were briefed about the experiment and notified that they would be blindfolded for the rest of the study. They were informed that Cross-Modal Box included an FDA approved intraoral interface, which might create tingling feelings on the tongue. In any discomfort, they were also informed that they have the right to withdraw from the experiment at any time. The study consisted of two phases: training and main experiment.

3.4.5.1 Training

Cross-Modal Box was introduced, and blindfolded participants were trained with the unisensory display modes separately. The order of display modes was also altered between participants. The trainings were not extensive and took less than 5 minutes to complete. Each training stimulus was either placed in Cross-Modal Box to experience the tactile mode or its auditory equivalent was played in a continuous loop to train with the auditory mode. Participants were then asked to identify the orientation of a line stimulus or count the number of dots. Feedback was provided and how an image was transformed with the cross-modal displays was explained. Participants were not trained with the multisensory mode. A short break was given at the end.

3.4.5.2 Experiment

The main experiment followed a similar procedure to the training phase. Prior to exploring a stimulus, blindfolded participants were verbally asked “what shape/letter is it?” and provided with four randomised options. They reported their responses verbally. The stimuli were always placed in Cross-Modal Box with respect to the viewpoint of the participants. No feedback was given.

For each trial with the tactile mode, participants were asked to place the intra-oral interface on their tongues once a stimulus was correctly placed in Cross-Modal Box. For each trial with the auditory mode, sonifications were played in a loop. For each trial with the multisensory mode, the auditory cues were played as soon as participants placed the intra-oral interface on their tongues. The tactile and/or auditory display modes were immediately stopped once the participants announced their response. Along with participant responses, reaction times between the initiation of the display mode and its stop were recorded in milliseconds. A short break was given between conditions.

Having completed the three conditions, participants were given a post-experiment questionnaire, which asked them to write down the strategies they used to identify objects and also rank the tasks and conditions in difficulty level.

3.5 Results

The primary objective of this experiment was to apply a framework of multisensory principles in investigating the benefits of multisensory as opposed to unisensory display modes (H1) and whether this happens in an optimal fashion (i.e. multisensory integration) (H2). For these evaluations, performance with respect to response accuracies and times were analysed between display mode (i.e. auditory, tactile and audio-tactile) and stimuli type (i.e. letters or shapes). Gender differences were also explored. Reliability was calculated for investigating the difficulty of display modes and

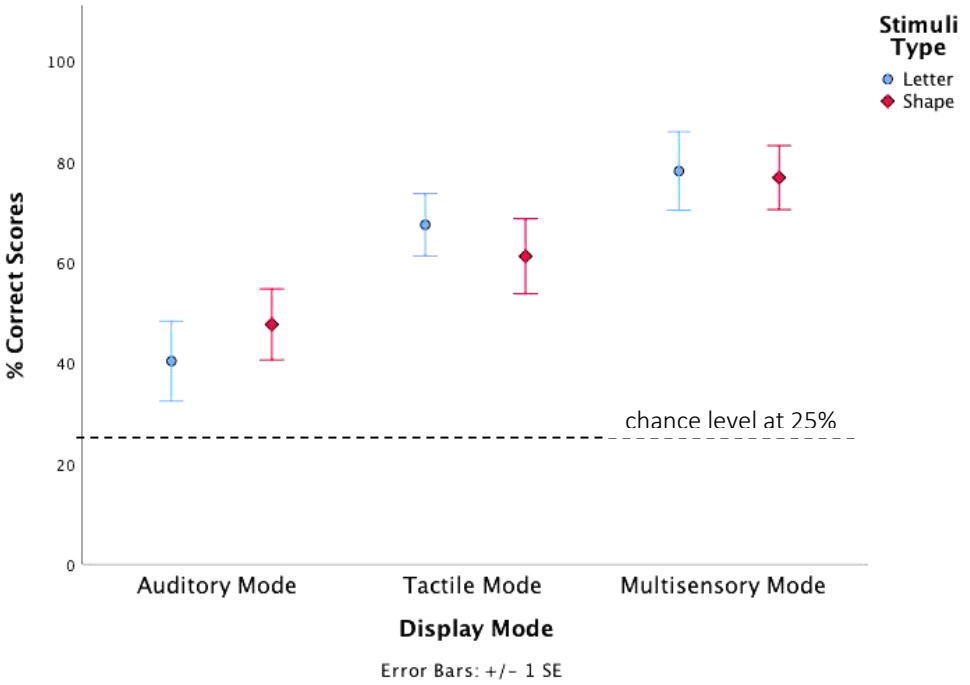
stimuli, and a thematic analysis was carried out for qualitative strategies used. The quantitative analysis was completed with SPSS25 [45] and the qualitative analysis was carried out with two coders using ATLAS.ti [5].

3.5.1 Calculating Response Accuracy and Reaction Times

In total, each participant completed three conditions of 16 object recognition trials (eight letters and eight shapes). Responses were assigned to binary scores ('1' for a correct response and '0' for an incorrect response) and then aggregated across stimulus types to calculate response accuracy in percentages for each display mode per participant. For each trial, reaction times were recorded in milliseconds. These were averaged for stimulus types across display modes per participant.

3.5.1.1 Analysing Response Accuracy

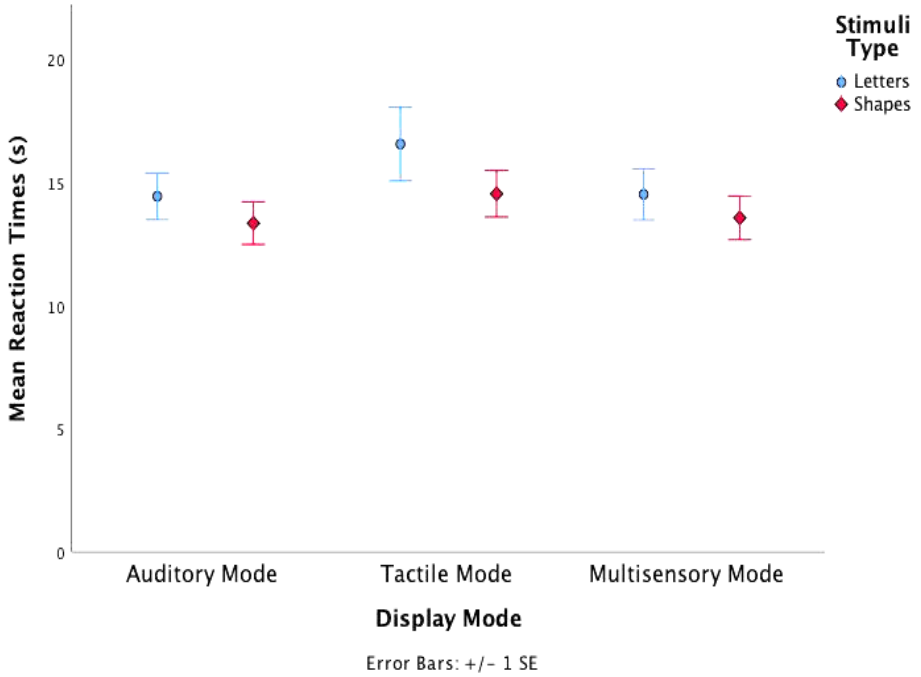
A statistically significant two way interaction was found between display mode and stimuli type, F(2,92) = 6.082, $p = .003$, partial $\eta^2 = 0.117$ (Graph 3.1). Overall, response accuracies were statistically different between the auditory mode (M = 44.0%, SE = 3.0), tactile mode (M = 64.3%, SE = 3.0), and multisensory mode (M = 77.5%, SE = 3.4), F(2,92) = 47.640, $p = <.001$, partial $\eta^2 = 0.509$. Bonferroni corrected pairwise comparisons confirmed these were statistically significant: with the mean differences of 20.3% (95% CI, [11.0, 29.6], $p<.001$) between tactile and auditory modes, 13.2% (95% CI, [6.5, 19.8], $p < .001$) between tactile and multisensory modes, and 33.5% (95% CI, [23.9, 43], $p < .001$) between auditory and multisensory modes. Response accuracies were not statistically different between letter (M = 62.0%, SE = 2.6) and shape (M = 61.9%, SE = 2.6) recognition, F(1,46) = 0.002, $p = .966$, partial $\eta^2 = 0.000$.



Graph 3.1 represents accuracy with respect to the display mode and stimuli type. Error bars show +/- 1 SE. The average correct scores for letter stimuli via auditory, tactile and multisensory modes were respectively 40.4% (SE = 4.0), 67.5% (SE = 3.1) and 78.1% (SE = 3.9). The average scores for shape stimuli were respectively 47.7% (SE = 3.5), 61.2% (SE = 3.7), 76.8% (SE = 3.2).

3.5.1.2 Reaction Times

No statistically significant two-way interaction was found between display mode and stimuli type, F(2,94) = 0.856, *p = .428, partial* η^2 *= 0.018* (Graph 3.2). The main effect of stimuli type showed that there was a statistically significant difference between shapes (M = 13.8s, SE = 0.81) and letters (M = 15.17s, SE = 1.06), F(1,47) = 6.249, *p = .016, partial* η^2 *= 0.117*, with a mean difference of 1.35s (95% CL, [0.264, 2.441]). The main effect of display mode also indicated a significant difference, F(2,94) = 3.541, *p = .033, partial* η^2 *= 0.070*; however, Bonferroni corrected pairwise comparisons did not find any statistical differences between auditory mode (M = 13.9s, SE = 0.86), tactile mode (M = 15.5s, SE = 1.17) and multisensory mode (M = 14.04s, SE = 0.89). Overall reaction times of 'correct' (M = 14.9s, SD = 10.2) and 'incorrect' (M = 14.2, SD = 11.1) responses were also analysed. This inspection did not find any significant differences, with a mean difference of 0.7s (95% CI, [-0.22, 1.6]), t(2302) = 1.488, *p = .137.*



Graph 3.2 shows the mean reaction times in seconds in relation to the display mode and stimuli type. Error bars show +/- 1 SE. The average reaction times for letter stimuli via auditory, tactile and multisensory modes were respectively 14.4s (SE = 1.0), 16.6s (SE = 1.5) and 14.5s (SE = 1.0). The average reaction times for shape stimuli were respectively 13.3s (SE = 0.9), 14.5s (SE = 0.9) and 13.6s (SE = 0.9).

3.5.2 Gender Differences

3.5.2.1 Response Accuracy

There was a statistically significant difference in response accuracies between males (M = 66.4%, SE = 3.4) and females (M = 57.5%, SE = 3.4), F(2,92) = 4.832, $p = .01$, *partial* $\eta^2 = 0.095$. Among display modes, only with the auditory display, a significant difference was found with a difference of 21.4% (95% CI, [11.6, 31.1]), t(94) = 4.36, $p < .001$ between males and females. Among stimuli type, only with shapes, a similar difference was found with a significant difference of 11.3% (95% CI [2.6, 19.9]), t(142) = 2.58, $p = .01$

3.5.2.2 Reaction Times

Even though a statistically significant interaction between gender and display mode was found, F(2,92) = 114.187, $p = .006$, *partial* $\eta^2 = 0.079$, Bonferroni corrected

contrasts did not point to a significant difference. Similarly, no significant difference was found between gender and stimuli type, F(1,46) = 9.805, *p = .5, partial* η^2 *= 0.010*.

3.5.3 Task and Display Mode Difficulty

3.5.3.1 Task Difficulty

To measure the agreement between participants' responses to task difficulty questions, Kendal's W was calculated. A statistically significant agreement within participants' responses was found, W = 0.396, p<.001 (Table 3.1).

Table 3.1 summarises the percentages of responses given to the task difficulty questions, where the elements in the first row were asked to be ordered from the easiest to the hardest for each element in the first column. Cells are shaded with respect to the value of percentages (i.e. 100% is pitch black and 0% is white).

	Overall Task Difficulty	Auditory Mode	Tactile Mode	Multisensory Mode
Recognising Letters	25.0%	10.4%	58.3%	12.5%
Recognising Shapes	75.0%	87.5%	25.0%	25.0%
Equally Easy	x	2.1%	16.7%	62.5%

3.5.3.2 Display Mode Difficulty

Similarly, Kendal's W was calculated to measure the agreement between participants' responses to display mode difficulty questions. A statistically significant agreement within participants' responses was found, W = 0.484, p <.001 (Table 3.2).

Table 3.2 summarises the percentages of responses given to the display mode difficulty questions, where participants were asked to order the display modes from the easiest to use to the hardest. Cells are shaded with respect to the value of percentages (i.e. 100% is pitch black and 0% is white).

	Auditory Mode	Tactile Mode	Multisensory Mode
Easy Difficulty	8.3%	66.7%	25.0%
Medium Difficulty	8.3%	29.2%	62.5%
Hard Difficulty	83.3%	4.2%	12.5%

3.5.4 Qualitative Strategies

3.5.4.1 Auditory Mode

Eighty-seven percent of participants reported that they identified the stimuli globally by focusing on the changes in pitch from left to right and the duration of the auditory stimuli. This made it possible to recognise the defined characteristics of the stimuli (e.g. “curvy”, “straight”, “diagonal”, “louder in the middle of the sound clip” and “horizontally elongated”). While these “defined features” often made some response options more distinct than others, helping participants “narrow down” their choices, 92% of participants struggled to identify local details. Participants found vertical lines in particular to be “too difficult” to recognise. Eighty-four percent of participants reported that they identified constant tones first in order to recognise horizontal lines and then looked for “gradual changes in pitch” for “diagonal lines”. This was because identifying a “left to right horizontal scan” for “changes in pitch” and a vertical shift in elevation made the recognition of slopes harder than horizontal lines. Ninety-four percent of participants also found the recognition of “lines” easier than “curvy shapes”. Overall, 85% of participants tried to “visualise” the overall soundscape and “relate it to the options”, eliminating the ones that did not match the soundscapes.

3.5.4.2 Tactile Mode

Ninety-three percent of participants reported that they initially identified the stimuli globally as a whole image. In this way, it was possible for them to imagine the “contours” of the tactile stimuli such as “defined edges”, “straight” or “curved lines”, “vertices”, “perimeter” and “symmetry”. Then, by either moving the intraoral interface or their tongues freely, 90% of participants identified local details. Identifying local details further entailed a “vertical motion” or a “top-bottom” scan of the stimuli (e.g. vertical lines). Thirty-seven percent of participants also mentioned that they performed a “horizontal” or “right to left” scan after the vertical motion to recognise “finer details”. Eighty-nine percent of participants stated that they tried to “create an image in my mind”, resembling the sensations from the tactile display mode to “touching the shapes with hand”. Contrarily, 8% of participants reported that

they used “process of elimination” to go over each response option, matching the electrotactile sensations to the options provided.

3.5.4.3 Multisensory Mode

Ninety-four percent of the participants reported that they used the multisensory cues to “confirm” or “reinforce” their responses. A few reported that they “exclusively” relied on the tactile mode and “ignored” the auditory mode as they were too “complicated” and “less precise”. Eighty-seven percent of participants also stated that they identified the “global” and “horizontal” features of the stimuli with the auditory mode and “finer details” and “vertical features” with the tactile mode. One participant specifically mentioned, “Initially I felt the combination made it more difficult to identify shapes/letters, so then I began to separate the two devices out, paying attention to the sound to get an idea of the shape and confirming details with my tongue”. Additionally, 92% of participants reported that recognising “horizontal lines” were easier with the auditory mode than the tactile mode.

3.6 Discussion

In the current research, a framework of multisensory principles was reviewed and then operationalised in an empirical study examining a cross-modal display prototype with unisensory and multisensory display modes. It was hypothesised that, following this framework, a multisensory display mode would outperform its unisensory components (H1) and this enhancement would be the result of multisensory integration (H2). In order to investigate these hypotheses, accuracy and reaction times were analysed and participants’ strategies were evaluated.

3.6.1 Summary of Results

The results indicated that the multisensory display mode of Cross-Modal Box with 77.5% accuracy significantly improved performance in comparison to the unisensory displays, with 44% accuracy in the auditory and 64.3% accuracy in the tactile modes. These findings were in line with H1.1. In comparing the unisensory display modes, the

tactile mode also resulted in significantly higher accuracy than the auditory mode, showing tactile superiority over the auditory display mode when the information capacity of cross-modal displays was equal. Moreover, accuracy did not significantly change between letters and shapes, suggesting the stimuli overall had similar task complexity. On the other hand, reactions times did not change significantly between the auditory (13.9s), tactile (15.5s) and multisensory (14.0s) display modes. This rejected H1.2 as the multisensory mode did not speed up reaction times. Given that the auditory mode resulted in the lowest accuracy yet the fastest reaction times, this could have rather been due to a trade-off between the accuracy and speed. Consequently, reaction times of correct and incorrect responses were further analysed. Despite correct responses were 0.7s slower than the incorrect ones, this difference was not significant. This confirmed that there was no such trade-off. Recognising shapes were significantly faster than recognising letters with a difference of 1.35s. This is in line with the fact that most participants found shape recognition easier. In investigating gender differences, it was also evidenced that males performed significantly better with the auditory display mode with a difference of 21.4%, and in recognising shapes with a difference of 11.3%. No difference between genders was observed in reaction times.

With the difficulty questions, participants reached a significant level of agreement. Overall, 75% of the participants agreed that recognising shapes was easier than recognising letters. 88% of participants found it easier to recognise shapes with the auditory mode and their accuracy was higher despite being insignificant to letter recognition. 58% of them found it easier to recognise letters with the tactile mode and their accuracy was higher despite being insignificant to shape recognition. 63% of them found both the stimuli type equally easy to recognise and there was no significant difference between shape and letter recognition. 83% of participants found the auditory mode the hardest display mode, which significantly resulted in the lowest overall score; 67% of them found the tactile mode the easiest display mode; and 63% found the multisensory mode at medium difficulty. These further confirmed similarities between participants' quantitative responses and their qualitative strategies.

Participants reported the auditory mode to be the most helpful in recognising global details of a stimulus, especially on the horizontal plane. In addition to providing the global features, with the tactile mode, participants reported that they were able to identify local details, especially on the vertical plane. Participants used the multisensory mode to complement their sensory judgements from the modality-specific cues, combining the horizontal and vertical resolution of the auditory and tactile display modes respectively.

3.6.2 Theoretical Implications

In examining Cross-Modal Box with unisensory and multisensory modes, it was shown the multisensory mode was superior in accuracy but not in reaction times. This rejected **H2.1.** On the other hand, participants' strategies reflected the benefits of multisensory processing, which was in line with **H2.2**. Their strategies also indicated that participants made conscious decisions, especially while using the multisensory display mode. Participants actively combined different channels of information (e.g. horizontal and global features from the auditory mode, and vertical and local details from the tactile mode) to construct a more robust perceptual judgement about the sensory source. These explain why the multisensory display mode resulted in the highest performance. Moreover, multisensory integration is conceptualised as an autonomous and unconscious process [100]. It was therefore concluded that enhanced performance in fact was not because of multisensory integration (**H2**).

The multisensory framework reviewed earlier suggests an alternative explanation to the experimental hypotheses and why the multisensory display mode led to higher accuracy but not faster reaction times. Participants' strategies evidenced that each unisensory display mode provided a set of complementary advantages than the other. For example, among one of the mutually emphasised strategic themes was the axis speciality of the auditory and tactile modes. Participants found the auditory display mode to have higher horizontal resolution, and the tactile mode to have higher vertical resolution. In using the multisensory mode, they combined these features to complement their sensory judgments, which resulted in higher accuracy. In line with

the multisensory framework, this suggested that participants actually benefited from multisensory combination and not integration. Having no differences in reaction times between display modes also meant that cross-modal displays could benefit from multisensory combination without a speed trade-off.

3.6.2.1 Limitations and Future Perspectives

In prototyping Cross-Modal Box, it was assumed that sensory substitution techniques would maintain the mutual characteristics of the sensory source. It was therefore concluded that the information conveyed via the auditory and tactile modes would be redundant when they were spatially and temporally aligned. The redundancy was expected because sensory substitution techniques would deliver the same information via different display modes. According to the principles of spatial and temporal coincidence, it was hypothesised that this would eventually lead to multisensory integration and respectively reduce reaction times while improving performance. Multisensory integration occurs when sensory information is redundant whereas this is not necessary for multisensory combination. Instead, the cross-modal display modes were used as complementary information channels in a way described in participants' strategies. Their strategies further indicated that the information delivered from the auditory and tactile display modes were not redundant. Instead, participants were able to pick different levels of information from each display mode. This eventually enabled them to utilise multisensory combination, which respectively improved their multisensory response quality while reaction times were not affected.

We therefore argue that the reason why the hypothesis that predicted multisensory integration was rejected was because of the assumption on redundancy of information delivered with sensory substitution techniques. This concluded that it was multisensory combination that improved the user responses. We further suggest that cross-modal displays do not always carry on the characteristics of the sensory origin. The question of which sensory experience sensory substitution belongs to has been long debated. The deference thesis claims that the sensory substitution experience switches to the substituted sense (e.g. 'seeing with the skin' [104], 'seeing with the

brain' [9], and 'seeing with sound' [68]) [44,71,73]. This thesis is in line with our assumption that deduced cross-modal feedback from auditory and tactile feedback would be redundant as they substituted for the same information source. On the other hand, the dominance theory argues that sensory substitution experience would remain in the substituting display mode [11,79]. In contrast to these polarising arguments, a recent line of research also proposes the vertical integration theory as the manifestation of the two [3,6,8,24,25]. This view evidences that cross-modal displays, powered with sensory substitution techniques like Cross-Modal Box, could enable pre-existing capacities of multiple senses for the given task and alter users' cognitive strategies accordingly.

Participants reported their strategies both in visual terms, such as "visualising", and also in distinct expressions specific to the display modes, such as the notions of horizontal and vertical resolutions. This suggested that the cross-modal display modes enabled pre-existing capacities of multiple senses, supporting the vertical integration theory. In the light of this thesis, it could then be argued that participants optimised their cognitive strategies with respect to the display modes. This is plausible because, despite carrying the same information capacity, each display mode was recognised to have a different kind of perceived resolution. The perceived resolution was further enhanced via the multisensory mode. That is, participants were able to use both of the unisensory display modes in a complimentary way via multisensory combination.

In order to investigate this further, we tentatively looked at whether the perceived resolution of the display modes influenced accuracy of recognising individual stimulus. For example, differentiating a square from a rectangle depends on understanding whether the height (vertical resolution) is equal to the width (horizontal resolution) of the shape. It could therefore be expected that the auditory mode with higher horizontal resolution would be biased towards a rectangle. Similarly, the tactile mode with the higher vertical resolution would be biased towards a square. In fact, response frequencies from the auditory mode showed that rectangles were correctly identified with 77.1% accuracy as opposed to mistaking it for a square with 10.4% of the time. Meanwhile, squares were recognised as rectangles 62.5% of the time as opposed to

their correct recognition of 27.1% accuracy. As expected, response frequencies from the tactile mode also revealed that squares were correctly identified with 54.2% accuracy as opposed to mistaking it for a rectangle with 27.1% of the time. Meanwhile, rectangles were recognised as squares with 45.8% of the time as opposed to their correct recognition of 37.5% accuracy. Moreover, it was observed that this was not the case with the multisensory mode. As such examination was beyond the scope of the current experimental design, this investigation was not carried further. However, the cross tabulation of stimuli type and response frequencies for each display mode can be found in Appendix 3.C. Future research should therefore target how different display modes influence the perceived resolution along with performance. This insight could further be conceptualised in maximising the benefits of multisensory combination such that the perceived multisensory resolution from a cross-modal display is enhanced.

3.6.3 Practical Implications

A recent line of research showed that enabling users to switch the display between tactile and visual modes in navigation applications could outperform unisensory modes [58]. Furthermore, such display mode alteration could be adaptive and achieved at algorithmic level via artificial intelligence for smart user experiences [57,59]. The findings from the current research also support that multisensory displays could outperform their unisensory counterparts and improve user experience via multisensory combination. By applying multisensory combination as a complementary technique to enrich the perceived resolution of multisensory modes, it could be possible to understand when various display modes are useful. In this way, users could benefit from multisensory systems in unisensory, alternating and multisensory modes depending on their cognitive strategies and needs for the given task. In parallel, intelligent algorithms that selectively adapts to alternating display modes could be improved with more insight from research targeting multisensory combination and perceived resolution. Additionally, tangible interactions could be developed further to complement and enhance how we interact with the digital world and extended reality platforms. Overall, the applications of multisensory combination would practically

reach to a wider range of users as well as expand the use cases of many unisensory display modes.

Improving the status quo of cross-modal displays inherently encourages the development of inclusive technologies. That is, with sensory substitution techniques, it is possible to prototype displays that are both useable and accessible to a wider user group. Training requirements of cross-modal displays with unisensory modes, however, have been blamed for poor user experience and scarce user adoption [61,62,76]. Participants in the current study were trained minimally with Cross-Modal Box. Nonetheless, their performance with the multisensory mode seems to outperform what previous studies, which investigated cross-modal displays with unisensory modes, reported. The current findings further suggested that poorer performance with unisensory cross-modal displays could be related to how the perceived resolution influence overall performance. In contrast to the deference and dominance theses, if cross-modal displays deliver the features of both the substituted and substituting sensory sources, then it means that unisensory display modes can only partially represent the sensory source. While it was evidenced that lengthier trainings were correlated with higher performance with cross-modal displays, this could as well be because increasing the training duration compensates for the missing resolution. Overall, studying cross-modal displays in relation to multisensory combination could minimise their training requirements, thereby widening their adoption in novel use cases, assistive and mainstream alike.

3.7 Conclusion

Applying a framework of multisensory processing principles to display development bridges the gap between multisensory research and HCI. The current research successfully demonstrated this by examining a cross-modal display prototype with unisensory and multisensory modes. It was evidenced that, instead of multisensory integration, the multisensory combination of cross-modal cues could enhance performance by complementing the perceived resolution of sensory sources. Despite

carrying the same information capacity, the auditory and tactile cross-modal cues delivered different perceived resolutions. While the auditory feedback was higher in horizontal resolution, the tactile feedback was higher in vertical resolution. The differences in axis specific resolutions were complemented by multisensory combination, thereby enhancing the performance from the multisensory mode. The current methodology highlighted the importance of a mixed study design in investigating this issue. Understanding participants' strategies is as much important as the quantitative data collected from their behavioural responses. Vertical integration theory was also supported by showing how sensory substitution techniques could be embodied both in the substituted and substituting senses. This was further evident in how the perceived multisensory resolution was enhanced via multisensory combination. Participants were able to strategize with respect to different display modes and achieved significantly higher performance via the multisensory mode without slowing down. Overall, the current research exemplifies how display development can be enhanced with multisensory combination techniques and conclusively presents future research directions for developing inclusive technologies.

3.8 References

1. Amir Amedi and Shlomo Hanassy. 2011. Infra red based devices for guiding blind and visually impaired persons. Retrieved from https://patents.google.com/patent/WO2012090114A1/en
2. Apple. 2019. Keynote. *Apple*. Retrieved from https://www.apple.com/keynote/
3. Gabriel Arnold, Jacques Pesnot-Lerousseau, and Malika Auvray. 2017. Individual Differences in Sensory Substitution. *Multisensory Research* 30, 6: 579–600. https://doi.org/10.1163/22134808-00002561
4. A Arnoldussen and D. C. Fletcher. 2012. Visual Perception for the Blind: The BrainPort Vision Device. *Retinal Physician* 9, 1: 32–34.
5. ATLAS.ti. 2019. ATLAS.ti: The Qualitative Data Analysis & Research Software. *ATLAS.ti*. Retrieved from https://atlasti.com/
6. Malika Auvray and Mirko Farina. 2017. Patrolling the Boundaries of Synaesthesia. In *Synaesthesia: Philosophical & Psychological Challenges*, O Deroy (ed.). Oxford University Press, Oxford, 248–274.
7. Malika Auvray, Sylvain Hanneton, and J Kevin O'Regan. 2007. Learning to Perceive with a Visuo — Auditory Substitution System: Localisation and Object Recognition with 'The Voice.' *Perception* 36, 3: 416–430. https://doi.org/10.1068/p5631
8. Malika Auvray and Erik Myin. 2009. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science* 33, 6: 1036–1058. https://doi.org/10.1111/j.1551-6709.2009.01040.x
9. Paul Bach-y-Rita and Stephen W. Kercel. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences* 7, 12: 541–546. https://doi.org/10.1016/J.TICS.2003.10.013
10. Fernando Bermejo, Ezequiel A. Di Paolo, Mercedes X. Hüg, and Claudia Arias. 2015. Sensorimotor strategies for recognizing geometrical shapes: a comparative study with different sensory substitution devices. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00679
11. Ned Block. 2007. Spatial Perception via Tactile Sensation. In *Consciousness, Function, and Representation*. The MIT Press. https://doi.org/10.7551/mitpress/2111.003.0020
12. Andrew J. Bremner. 2017. Multisensory Development: Calibrating a Coherent Sensory Milieu in Early Life. *Current Biology* 27, 8: R305–R307. https://doi.org/10.1016/J.CUB.2017.02.055
13. David J. Brown, Andrew J. R. Simpson, and Michael J. Proulx. 2014. Visual Objects in the Auditory System in Sensory Substitution: How Much Information Do We Need? *Multisensory Research* 27, 5–6: 337–357. https://doi.org/10.1163/22134808-00002462
14. Galit Buchs, Shachar Maidenbaum, and Amir Amedi. 2014. Obstacle Identification and Avoidance Using the 'EyeCane': a Tactile Sensory Substitution Device for Blind Individuals. . Springer, Berlin, Heidelberg, 96–103. https://doi.org/10.1007/978-3-662-44196-1_13
15. Galit Buchs, Shachar Maidenbaum, and Amir Amedi. 2015. Augmented non-visual distance sensing with the EyeCane. In *Proceedings of the 6th Augmented Human International Conference on - AH '15*, 209–210. https://doi.org/10.1145/2735711.2735780
16. Heinrich H. Bülthoff and Hanspeter A. Mallot. 1988. Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A* 5, 10: 1749. https://doi.org/10.1364/JOSAA.5.001749
17. Austin McRae Butts. 2015. Enhancing the Perception of Speech Indexical Properties of Cochlear Implants through Sensory Substitution. Arizona State University.
18. Gemma A. Calvert, Charles Spence, and Barry E. Stein. 2004. The Handbook of Multisensory Processing.
19. Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, 505–514. https://doi.org/10.1145/2501988.2502018
20. Jason S Chan, Corrina Maguinness, Danuta Lisiecka, Annalisa Setti, and Fiona N Newell. 2012. Evidence for Crossmodal Interactions across Depth on Target Localisation Performance in a Spatial Array. *Perception* 41, 7: 757–773. https://doi.org/10.1068/p7230
21. Yuri P. Danilov, Mitchell E. Tyler, and Kurt A. Kaczmarek. 2008. Vestibular sensory substitution using tongue electrotactile display. In *Human Haptic Perception: Basics and Applications*.

Birkhäuser Basel, Basel, 467–480. https://doi.org/10.1007/978-3-7643-7612-3_39
22. Yuri Danilov and Mitchell Tyler. 2005. BrainPort: An Alternative Input to the Brain. *Journal of Integrative Neuroscience* 04, 04: 537–550. https://doi.org/10.1142/S0219635205000914
23. David Eagleman. 2015. Can we create new senses for humans? *TED*. Retrieved from https://www.ted.com/talks/david_eagleman_can_we_create_new_senses_for_humans
24. Ophelia Deroy and Malika Auvray. 2012. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00457
25. Ophelia Deroy and Malika Auvray. 2014. A Crossmodal Perspective on Sensory Substitution. In *Perception and Its Modalities*. Oxford University Press, 327–349. https://doi.org/10.1093/acprof:oso/9780199832798.003.0014
26. Jon Driver and Charles Spence. 1998. Crossmodal attention. *Current Opinion in Neurobiology* 8, 2: 245–253. https://doi.org/10.1016/S0959-4388(98)80147-5
27. N.J. Droessler, D.K. Hall, M.E. Tyler, and N.J. Ferrier. 2001. Tongue-based electrotactile feedback to perceive objects grasped by a robotic manipulator: preliminary results. In *Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1404–1407. https://doi.org/10.1109/IEMBS.2001.1020464
28. Marc O. Ernst and Martin S. Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 6870: 429–433. https://doi.org/10.1038/415429a
29. Marc O. Ernst and Heinrich H. Bülthoff. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8, 4: 162–169. https://doi.org/10.1016/J.TICS.2004.02.002
30. Tayfun Esenkaya and Michael J. Proulx. 2016. Crossmodal processing and sensory substitution: Is “seeing” with sound and touch a form of perception or cognition? *Behavioral and Brain Sciences* 39: e241. https://doi.org/10.1017/S0140525X1500268X
31. Daniel J. Finnegan, Eamonn O’Neill, and Michael J. Proulx. 2016. Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 200–212. https://doi.org/10.1145/2858036.2858065
32. Claude Fortin, Kate Hennessy, Ruedi Baur, and Pierre Fortin. 2013. Beyond the vision paradigm. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays - PerDis ’13*, 91. https://doi.org/10.1145/2491568.2491588
33. M. A. Frens and A. J. Van Opstal. 1995. A quantitative study of auditory-evoked saccadic eye movements in two dimensions. *Experimental Brain Research* 107, 1: 103–117. https://doi.org/10.1007/BF00228022
34. Asif A. Ghazanfar and Charles E. Schroeder. 2006. Is neocortex essentially multisensory? *Trends in Cognitive Sciences* 10, 6: 278–285. https://doi.org/10.1016/J.TICS.2006.04.008
35. Stan C. A. M. Gielen, Richard A. Schmidt, and Pieter J. M. Van Den Heuvel. 1983. On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics* 34, 2: 161–168. https://doi.org/10.3758/BF03211343
36. Patricia Grant, Lindsey Spencer, Aimee Arnoldussen, Rich Hogle, Amy Nau, Janet Szlyk, Jonathan Nussdorf, Donald C. Fletcher, Keith Gordon, and William Seiple. 2016. The Functional Performance of the BrainPort V100 Device in Persons who Are Profoundly Blind. *Journal of Visual Impairment & Blindness* 110, 2: 77–88. https://doi.org/10.1177/0145482X1611000202
37. Alastair Haigh, David J. Brown, Peter Meijer, and Michael J. Proulx. 2013. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00330
38. Giles Hamilton-Fletcher, Marianna Obrist, Phil Watten, Michele Mengucci, and Jamie Ward. 2016. "I Always Wanted to See the Night Sky": Blind User preferences for Sensory Substitution Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 2162–2174. https://doi.org/10.1145/2858036.2858241
39. Maurice Hershenson. 1962. Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology* 63, 3: 289–293. https://doi.org/10.1037/h0039516
40. Rebekka Hoffmann, Simone Spagnol, Árni Kristjánsson, and Runar Unnthorsson. 2018. Evaluation of an Audio-haptic Sensory Substitution Device for Enhancing Spatial Awareness for the Visually Impaired. *Optometry and Vision Science* 95, 9: 757–765. https://doi.org/10.1097/OPX.0000000000001284
41. E.E. Hoggan and S.A. Brewster. 2006. Crossmodal Interaction with Mobile Devices. In *Visual*

Languages and Human-Centric Computing (VL/HCC'06), 234–235. https://doi.org/10.1109/VLHCC.2006.18

42. Eve Hoggan and Stephen Brewster. 2007. Designing audio and tactile crossmodal icons for mobile devices. In *Proceedings of the ninth international conference on Multimodal interfaces - ICMI '07*, 162. https://doi.org/10.1145/1322192.1322222
43. Eve Hoggan, Topi Kaaresoja, Pauli Laitinen, and Stephen Brewster. 2008. Crossmodal congruence. In *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*, 157. https://doi.org/10.1145/1452392.1452423
44. Susan Hurley and Alva Noë. 2003. Neural Plasticity and Consciousness. *Biology & Philosophy* 18, 1: 131–168. https://doi.org/10.1023/A:1023308401356
45. IBM. 2019. SPSS Software. *IBM*. Retrieved from https://www.ibm.com/analytics/spss-statistics-software
46. Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, 234–241. https://doi.org/10.1145/258549.258715
47. Ali Israr, Olivier Bau, Seung-Chan Kim, and Ivan Poupyrev. 2012. Tactile feedback on flat surfaces for the visually impaired. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 1571. https://doi.org/10.1145/2212776.2223674
48. H. Jacobson. 1950. The Informational Capacity of the Human Ear. *Science* 112, 2901: 143–144. https://doi.org/10.1126/science.112.2901.143
49. H. Jacobson. 1951. The Informational Capacity of the Human Eye. *Science* 113, 2933: 292–293. https://doi.org/10.1126/science.113.2933.292
50. Myounghoon Jeon. 2015. Auditory User Interface Design: Practical Evaluation Methods and Design Process Case Studies. *The International Journal of Design in Society* 8, 2: 1–16. https://doi.org/10.18848/2325-1328/CGP/v08i02/38527
51. Ómar Jóhannesson, Oana Balan, Runar Unnthorsson, Alin Moldoveanu, and Árni Kristjánsson. 2016. The Sound of Vision Project: On the Feasibility of an Audio-Haptic Representation of the Environment, for the Visually Impaired. *Brain Sciences* 6, 3: 20. https://doi.org/10.3390/brainsci6030020
52. K.A. Kaczmarek, J.G. Webster, P. Bach-y-Rita, and W.J. Tompkins. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38, 1: 1–16. https://doi.org/10.1109/10.68204
53. Maria Karam, Carmen Branje, Gabe Nespoli, Norma Thompson, Frank A. Russo, and Deborah I. Fels. 2010. The emoti-chair. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, 3069. https://doi.org/10.1145/1753846.1753919
54. Christoph Kayser and Nikos K. Logothetis. 2007. Do early sensory cortices integrate cross-modal information? *Brain Structure and Function* 212, 2: 121–132. https://doi.org/10.1007/s00429-007-0154-0
55. R M Klein. 1977. Attention and visual dominance: a chronometric analysis. *Journal of experimental psychology. Human perception and performance* 3, 3: 365–78. https://doi.org/10.1037//0096-1523.3.3.365
56. Kenneth Kokjer. 1987. The Information Capacity of the Human Fingertip. *IEEE Transactions on Systems, Man, and Cybernetics* 17, 1: 100–102. https://doi.org/10.1109/TSMC.1987.289337
57. Kyle Kotowick and Julie Shah. 2017. Intelligent Sensory Modality Selection for Electronic Supportive Devices. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, 55–66. https://doi.org/10.1145/3025171.3025228
58. Kyle Kotowick and Julie Shah. 2018. Modality Switching for Mitigation of Sensory Adaptation and Habituation in Personal Navigation Systems. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*, 115–127. https://doi.org/10.1145/3172944.3172980
59. Kyle Kotowick and Julie Shah. 2018. Effects of an Adaptive Modality Selection Algorithm for Navigation Systems. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, 543–556. https://doi.org/10.1145/3242587.3242610
60. C. Lenay, S. Canu, and P. Villon. 1997. Technology and perception: the contribution of sensory substitution systems. In *Proceedings Second International Conference on Cognitive Technology*

Humanizing the Information Age, 44–53. https://doi.org/10.1109/CT.1997.617681
61. J. M. Loomis. 2010. Sensory substitution for orientation and mobility: what progress are we making? In *Foundations of Orientation and Mobility* (3rd ed.), William R. Wiener, Richard L. Welsh and Bruce B. Blasch (eds.). AFB Press, NewYork, 3–44.
62. Shachar Maidenbaum and Sami Abboud. 2014. Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews* 41: 3–15. https://doi.org/10.1016/J.NEUBIOREV.2013.11.007
63. Shachar Maidenbaum, Roni Arbel, Galit Buchs, Shani Shapira, and Amir Amedi. 2014. Vision through other senses: Practical use of Sensory Substitution devices as assistive technology for visual rehabilitation. In *22nd Mediterranean Conference on Control and Automation*, 182–187. https://doi.org/10.1109/MED.2014.6961368
64. Shachar Maidenbaum, Shelly Levy-Tzedek, Daniel-Robert Chebat, and Amir Amedi. 2013. Increasing Accessibility to the Blind of Virtual Environments, Using a Virtual Mobility Aid Based On the "EyeCane": Feasibility Study. *PLoS ONE* 8, 8: e72555. https://doi.org/10.1371/journal.pone.0072555
65. Shachar Maidenbaum, Shelly Levy-Tzedek, Daniel Robert Chebat, Rinat Namer-Furstenberg, and Amir Amedi. 2014. The Effect of Extended Sensory Range via the EyeCane Sensory Substitution Device on the Characteristics of Visionless Virtual Navigation. *Multisensory Research* 27, 5–6: 379–397. https://doi.org/10.1163/22134808-00002463
66. John J. McDonald, Wolfgang A. Teder-Sälejärvi, and Steven A. Hillyard. 2000. Involuntary orienting to sound improves visual perception. *Nature* 407, 6806: 906–908. https://doi.org/10.1038/35038085
67. P.B.L. Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 2: 112–121. https://doi.org/10.1109/10.121642
68. Peter Meijer. 2019. Seeing With Sound. Retrieved from https://www.seeingwithsound.com/
69. National Research Council. 2008. *Emerging Cognitive Neuroscience and Related Technologies*. National Academies Press, Washington, D.C. https://doi.org/10.17226/12177
70. Fiona N. Newell, Marc O. Ernst, Bosco S. Tjan, and Heinrich H. Bülthoff. 2001. Viewpoint Dependence in Visual and Haptic Object Recognition. *Psychological Science* 12, 1: 37–42. https://doi.org/10.1111/1467-9280.00307
71. Alva. Noë. 2004. *Action in perception*. MIT Press.
72. Scott D. Novich and David M. Eagleman. 2014. A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In *2014 IEEE Haptics Symposium (HAPTICS)*, 1–1. https://doi.org/10.1109/HAPTICS.2014.6775558
73. J. K. O'Regan. 2011. *Why red doesn't sound like a bell : understanding the feel of consciousness*. Oxford University Press.
74. Marianna Obrist, Elia Gatti, Emanuela Maggioni, Chi Thanh Vi, and Carlos Velasco. 2017. Multisensory Experiences in HCI. *IEEE MultiMedia* 24, 2: 9–13. https://doi.org/10.1109/MMUL.2017.33
75. Sharon Oviatt and Sharon. 1999. Ten myths of multimodal interaction. *Communications of the ACM* 42, 11: 74–81. https://doi.org/10.1145/319382.319398
76. Achille Pasqualotto and Tayfun Esenkaya. 2016. Sensory Substitution: The Spatial Updating of Auditory Scenes "Mimics" the Spatial Updating of Visual Scenes. *Frontiers in Behavioral Neuroscience* 10. https://doi.org/10.3389/fnbeh.2016.00079
77. Diane Pecher, René Zeelenberg, and Lawrence W. Barsalou. 2003. Verifying Different-Modality Properties for Concepts Produces Switching Costs. *Psychological Science* 14, 2: 119–124. https://doi.org/10.1111/1467-9280.t01-1-01429
78. Thomas J. Perrault, J. William Vaughan, Barry E. Stein, and Mark T. Wallace. 2003. Neuron-Specific Response Characteristics Predict the Magnitude of Multisensory Integration. *Journal of Neurophysiology* 90, 6: 4022–4026. https://doi.org/10.1152/jn.00494.2003
79. Jesse J. Prinz. 2006. Putting the brakes on enactive perception. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 12.
80. Michael J. Proulx, David J. Brown, Achille Pasqualotto, and Peter Meijer. 2014. Multisensory perceptual learning and sensory substitution. *Neuroscience & Biobehavioral Reviews* 41: 16–25. https://doi.org/10.1016/J.NEUBIOREV.2012.11.017
81. Michael J. Proulx and A. Harder. 2008. Sensory substitution: Visual-to-auditory sensory substitution devices for the blind. *Tijdschrift voor Ergonomie* 6, 33.

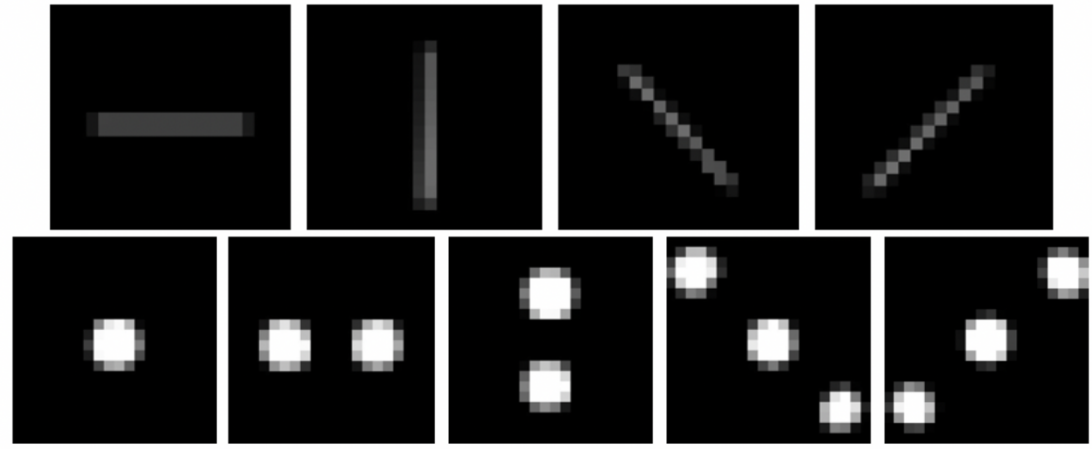
82. D. Rocchesso, R. Bresin, and M. Fernstrom. 2003. Sounding objects. *IEEE Multimedia* 10, 2: 42–52. https://doi.org/10.1109/MMUL.2003.1195160
83. I. Rock and J. Victor. 1964. Vision and Touch: An Experimentally Created Conflict between the Two Senses. *Science* 143, 3606: 594–596. https://doi.org/10.1126/science.143.3606.594
84. Marieke Rohde, Loes C J van Dam, and Marc Ernst. 2016. Statistically Optimal Multisensory Cue Integration: A Practical Tutorial. *Multisensory research* 29, 4–5: 279–317.
85. Hasti Seifi, Farimah Fazlollahi, Michael Oppermann, John Andrew Sastrillo, Jessica Ip, Ashutosh Agrawal, Gunhyuk Park, Katherine J. Kuchenbecker, and Karon E. MacLean. 2019. Haptipedia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–12. https://doi.org/10.1145/3290605.3300788
86. Charles Spence, Michael E. R. Nicholls, and Jon Driver. 2001. The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics* 63, 2: 330–336. https://doi.org/10.3758/BF03194473
87. Charles Spence, Marianna Obrist, Carlos Velasco, and Nimesha Ranasinghe. 2017. Digitizing the chemical senses: Possibilities & pitfalls. *International Journal of Human-Computer Studies* 107: 62–74. https://doi.org/10.1016/J.IJHCS.2017.06.003
88. Sharmila Sreetharan and Michael Schutz. 2019. Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope. *Multimodal Technologies and Interaction* 3, 1: 4. https://doi.org/10.3390/mti3010004
89. T. R. Stanford. 2005. Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience* 25, 28: 6499–6508. https://doi.org/10.1523/JNEUROSCI.5095-04.2005
90. Terrence R. Stanford and Barry E. Stein. 2007. Superadditivity in multisensory integration: putting the computation in context. *NeuroReport* 18, 8: 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315
91. B E Stein and M T Wallace. 1996. Comparisons of cross-modality integration in midbrain and cortex. *Progress in brain research* 112: 289–99. https://doi.org/10.1016/s0079-6123(08)63336-1
92. Barry E. Stein. 1998. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research* 123, 1–2: 124–135. https://doi.org/10.1007/s002210050553
93. Barry E. Stein, David Burr, Christos Constantinidis, Paul J. Laurienti, M. Alex Meredith, Thomas J. Perrault, Ramnarayan Ramachandran, Brigitte Röder, Benjamin A. Rowland, K. Sathian, Charles E. Schroeder, Ladan Shams, Terrence R. Stanford, Mark T. Wallace, Liping Yu, and David J. Lewkowicz. 2010. Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience* 31, 10: 1713–1720. https://doi.org/10.1111/j.1460-9568.2010.07206.x
94. Barry E. Stein and M. Alex. Meredith. 1993. *The merging of the senses*. MIT Press.
95. Barry E. Stein, W. Scott Huneycutt, and M. Alex Meredith. 1988. Neurons and behavior: the same rules of multisensory integration apply. *Brain Research* 448, 2: 355–358. https://doi.org/10.1016/0006-8993(88)91276-0
96. Noelle R. B. Stiles and Shinsuke Shimojo. 2015. Auditory Sensory Substitution is Intuitive and Automatic with Texture Stimuli. *Scientific Reports* 5, 1: 15628. https://doi.org/10.1038/srep15628
97. Noelle R. B. Stiles, Yuqian Zheng, and Shinsuke Shimojo. 2015. Length and orientation constancy learning in 2-dimensions with auditory sensory substitution: the importance of self-initiated movement. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00842
98. Ella Striem-Amit, Laurent Cohen, Stanislas Dehaene, and Amir Amedi. 2012. Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron* 76, 3: 640–652. https://doi.org/10.1016/J.NEURON.2012.08.026
99. H. Christiaan Stronks, Ellen B. Mitchell, Amy C. Nau, and Nick Barnes. 2016. Visual task performance in the blind with the BrainPort V100 Vision Aid. *Expert Review of Medical Devices* 13, 10: 919–931. https://doi.org/10.1080/17434440.2016.1237287
100. Durk Talsma, Daniel Senkowski, Salvador Soto-Faraco, and Marty G. Woldorff. 2010. The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences* 14, 9: 400–410. https://doi.org/10.1016/J.TICS.2010.06.008
101. M. Iftekhar Tanveer, A. S. M. Iftekhar Anam, Mohammed Yeasin, and Majid Khan. 2013. Do you

see what I see? In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '13*, 1–8. https://doi.org/10.1145/2513383.2513438

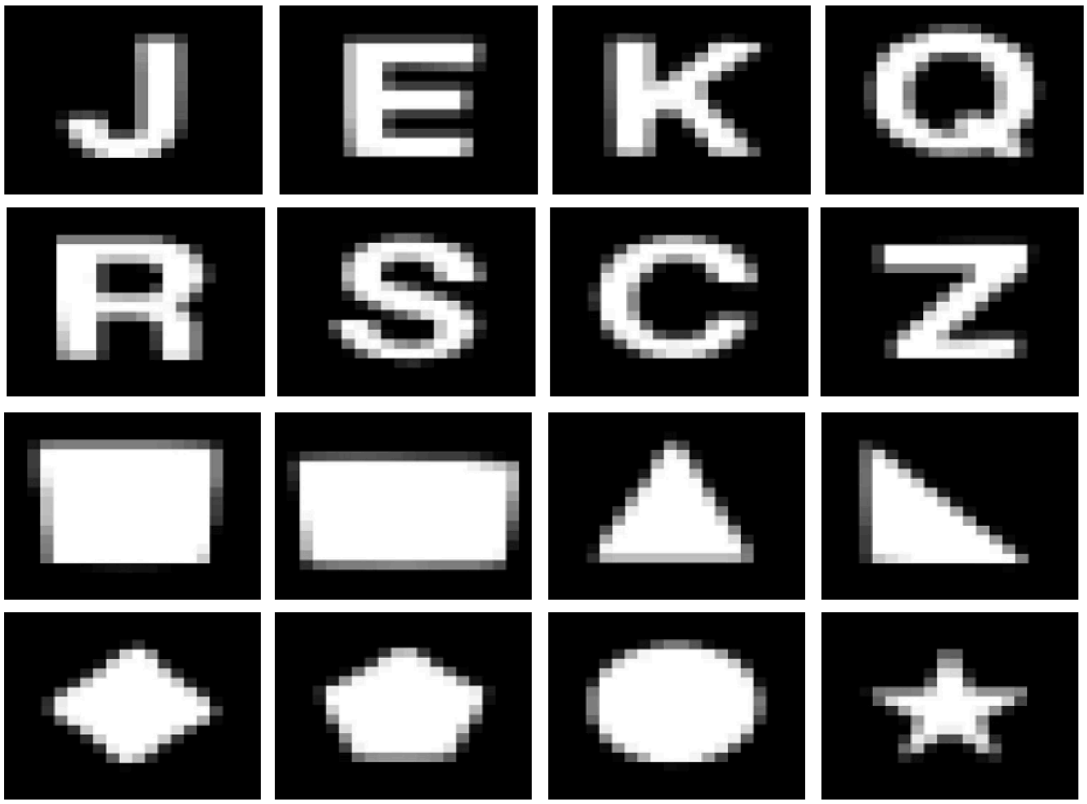
102. Jean Vroomen and Beatrice de Gelder. 2000. Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance* 26, 5: 1583–1590. https://doi.org/10.1037/0096-1523.26.5.1583
103. Jamie Ward and Peter Meijer. 2010. Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition* 19, 1: 492–500. https://doi.org/10.1016/J.CONCOG.2009.10.006
104. Benjamin W. White, Frank A. Saunders, Lawrence Scadden, Paul Bach-Y-Rita, and Carter C. Collins. 1970. Seeing with the skin. *Perception & Psychophysics* 7, 1: 23–27. https://doi.org/10.3758/BF03210126
105. Wicab. Wicab, Inc. | United States | BrainPort Technologies. Retrieved from https://www.wicab.com/wicab-inc

3.9 Appendix

Appendix 3.A: Training Stimuli



Appendix 3.B: Experimental Stimuli



Appendix 3.C: Cross Tabulation of Response Frequencies and Stimuli

The cross tabulations below represent the percentages of responses (in rows) given to stimuli (in columns). Cells are shaded with respect to the value of percentages (i.e. 100% is pitch black and 0% is white).

3.C.1 Letter Recognition with Auditory Mode

Responses (row) / Stimuli (column)	K	C	E	J	Q	S	R	Z
K	52.1%	12.5%	8.3%	27.1%	x	x	x	x
C	x	20.8%	x	x	29.2%	16.7%	x	33.3%
E	8.3%	x	50.0%	x	x	x	6.3%	35.4%
J	22.9%	x	12.5%	47.9%	x	x	16.7%	x
Q	x	x	x	6.3%	25.0%	45.8%	22.9%	x
S	14.6%	43.8%	x	x	12.5%	29.2%	x	x
R	x	31.3%	6.3%	x	x	12.5%	50%	x
Z	16.7%	18.8%	x	x	x	x	16.7%	47.9%

3.C.2 Shape Recognition with Auditory Mode

Responses (row) / Stimuli (column)	Square	Circle	Star	Rectangle	Right Triangle	Diamond	Triangle	Pentagon
Square	27.1%	4.2%	x	62.5%	x	x	6.3%	x
Circle	50.0%	35.4%	10.4%	x	x	x	4.2%	x
Star	x	6.3%	66.7%	x	x	x	2.1%	25.0%
Rectangle	10.4%	x	x	77.1%	10.4%	2.1%	x	x
Right Triangle	8.3%	4.2%	x	x	79.2%	8.3%	x	x
Diamond	x	x	x	10.4%	x	29.2%	43.8%	16.7%
Triangle	4.2%	41.7%	x	x	x	x	29.2%	25.0%
Pentagon	12.5%	31.3%	x	x	x	x	18.8%	37.5%

3.C.3 Letter Recognition with Tactile Mode

Responses (row) / Stimuli (column)	K	C	E	J	Q	S	R	Z
K	85.4%	4.2%	6.3%	4.2%	x	x	x	x
C	x	47.9%	x	x	39.6%	8.3%	x	4.2%
E	4.2%	x	68.8%	x	x	x	18.8%	8.3%
J	4.2%	x	4.2%	85.4%	x	x	6.3%	x
Q	x	x	x	8.3%	54.2%	35.4%	2.1%	x
S	2.1%	10.4%	x	x	43.8%	43.8%	x	x
R	x	6.3%	6.3%	x	x	4.2%	83.3%	x
Z	12.5%	x	x	x	x	x	16.7%	70.8%

3.C.4 Shape Recognition with Tactile Mode

Responses (row) / Stimuli (column)	Square	Circle	Star	Rectangle	Right Triangle	Diamond	Triangle	Pentagon
Square	54.2%	14.6%	x	27.1%	x	x	4.2%	x
Circle	20.8%	62.5%	8.3%	x	x	x	8.3%	x
Star	x	10.4%	43.8%	x	x	x	39.6%	6.3%
Rectangle	45.8%	x	x	37.5%	12.5%	4.2%	x	x
Right Triangle	0%	4.2%	x	x	87.5%	8.3%	x	x
Diamond	x	x	x	8.3%	x	77.1%	4.2%	10.4%
Triangle	4.2%	2.1%	x	x	x	x	85.4%	8.3%
Pentagon	14.6%	31.3%	x	x	x	x	12.5%	41.7%

3.C.5 Letter Recognition with Multisensory Mode

Responses (row) / Stimuli (column)	K	C	E	J	Q	S	R	Z
K	83.3%	0%	8.3%	8.3%	x	x	x	x
C	x	66.7%	x	x	16.7%	8.3%	x	8.3%
E	2.1%	x	72.9%	x	x	x	8.3%	16.7%
J	2.1%	x	2.1%	91.7%	x	x	4.2%	x
Q	x	x	x	0%	79.2%	6.3%	14.6%	x
S	4.2%	2.1%	x	x	16.7%	77.1%	x	x
R	x	4.2%	2.1%	x	x	8.3%	85.4%	x
Z	8.3%	6.3%	x	x	x	x	16.7%	68.8%

3.C.6 Shape Recognition with Multisensory Mode

Responses (row) / Stimulus (column)	Square	Circle	Star	Rectangle	Right Triangle	Diamond	Triangle	Pentagon
Square	68.8%	0%	x	29.2%	x	x	2.1%	x
Circle	18.8%	70.8%	6.3%	x	x	x	4.2%	x
Star	x	4.2%	83.3%	x	x	x	6.3%	6.3%
Rectangle	27.1%	x	x	70.8%	2.1%	0%	x	x
Right Triangle	0%	0%	x	x	95.8%	4.2%	x	x
Diamond	x	x	x	4.2%	x	83.3%	2.1%	10.4%
Triangle	2.1%	8.3%	x	x	x	x	81.3%	8.3%
Pentagon	0%	29.2%	x	x	x	x	10.4%	60.4%

Reflections IV

"I call our world Flatland, not because we call it so,
but to make its nature clearer to you, my happy readers,
who are privileged to live in Space."
Edwin A. Abbott[2]

The previous chapter demonstrated that the multisensory combination of auditory (sonifications) and tactile cross-modal display modes could enhance object recognition performance with cross-modal displays. This was explained by how perception with cross-modal displays might fall between the substituting and substituted senses. The results therefore further supported the vertical integration thesis and contributed to it in a multisensory context. The dimension-specific resolution of cross-modal displays, which the results of Chapter II tentatively pointed towards, was more evident in the previous study. That is, the tactile cross-modal feedback was reported to have higher perceived vertical resolution and the auditory cross-modal feedback had higher perceived horizontal resolution. It was concluded that the multisensory mode of Cross-Modal Box might be able to complement the axis-specific resolution for enhanced performance. A tentative comparison of the perception of 'square' and 'rectangle' was also presented, which lends preliminary support to this axis-specificity. Future research should further investigate this axis-specificity.

The previous study also supported the idea that multisensory displays could improve how we interact with the digital world. It was suggested that multisensory combination could be a practical method for developing inclusive displays with multisensory modes. If the vertical integration thesis holds true, multisensory combination could be utilised to supplement the dual nature of perception with cross-modal displays. In this way, inclusion could still be achieved by deploying a combination of cross-modal display modes so that a wider range of users could

[2] From Flatland: A Romance of Many Dimensions.

customise the sensory channels that work the best for them in different circumstances. Furthermore, the previous chapter highlighted the importance of semantic consistencies between theoretical and applied research, and the value of user strategies in interpreting results. In this respect, it offers a framework of multisensory processing principles for the HCI research and demonstrates how this could be applied as a methodology in multisensory display development.

The upcoming chapter will utilise the framework of multisensory processing and a similar methodology with the evaluation of two cross-modal display prototypes with unisensory and multisensory modes. By doing so, it will examine the inclusive applications of multisensory combination and the axis-speciality of cross-modal display modes in a navigation context. Accordingly, the next study further expands the inspection of perception with cross-modal displays in a multisensory context.

CHAPTER IV

Auditory and Tactile Cross-Modal Displays Can Help with the Last 10 Metres of Navigation

Declaration

This declaration concerns the article entitled:									
Auditory and Tactile Cross-Modal Displays Can Help with the Last 10 Metres of Navigation									
Publication status (tick one)									
Draft manuscript	x	Submitted		In review		Accepted		Published	
Publication details (reference)	N/A								
Copyright status (tick the appropriate statement)									
I hold the copyright for this material	x	Copyright is retained by the publisher, but I have been given permission to replicate the material here							
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	The candidate predominantly executed Formulation of ideas:80% Design of methodology:80% Experimental work:70% Presentation of data in journal format:90% For details, please see acknowledgements on the next page.								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed				**Date**					

Acknowledgements

I am thankful to Crescent Jicol and Simon Lange-Smith for their help with design, data collection and analysis of the research. I am grateful to Michael Proulx for his discussion of the research. I am also grateful to Vanessa Lloyd-Esenkaya for patiently proof-reading the manuscript. I also thank all the participants who took part in the experiment.

4.0 Abstract

Visual information, from physical traffic signs to digital web mapping services, dominates our sensory sources to aid successful navigation. While most users rely on graphical interfaces for navigation, there are inherent disadvantages to display modes that are predominantly visual. For those who can directly interact with visual interfaces (e.g. sighted users), being heavily dependent on a unisensory display mode prevents them from taking full advantage of their multisensory capabilities. For those who have limited access to graphical interfaces (e.g. visually impaired users), the inability to interact with popular mainstream applications obstructs equal participation in society. These issues formulate the basis of usability and accessibility challenges facing unisensory display modes in the context of developing inclusive navigation technologies. Additionally, the current navigation applications are not always capable of accommodating all the necessary perceptual information efficiently when the environment is dynamically changing. One additional challenge this issue creates is the development of display modes that assist users to finalise their journeys. The current navigation applications are good with providing feedback at macro-level (e.g. overall directions). Difficulties arise, however, in guiding users at micro-level navigation (e.g. pinpointing the target destination in the last 10-metres of navigation). An alternative way to tackle these challenges altogether is to develop displays that can optimise the information exchange between display modes and the user by facilitating non-visual cross-modal cues. To explore this alternative in the current research, two cross-modal displays, which can deliver spatial information in unisensory and multisensory modes (sonification and/or kinaesthesia, and sonification and/or tactile), were prototyped. In two experiments, the performance of these prototypes in complex micro-level navigation tasks were evaluated. The findings revealed that positional, two-dimensional tactile cues resulted in the highest performance. The two experiments also demonstrated how users accommodate various navigation strategies while using cross-modal displays. Overall, these findings can be applied in developing inclusive micro-level navigation applications.

4.1 Introduction

Assisted navigation became ubiquitous with the rise of smartphones as millions of users now seamlessly navigate physical environments with palm-sized, GPS-capable computers. Modern smartphones utilise Assisted GPS (A-GPS) technology that uses phone networks and GPS antennae to rapidly determine position [104]. This enables general guidance with a reasonable accuracy over large areas, equipping users with good macro-level navigation skills. Mainstream navigation applications (e.g. Apple Maps) typically interact with their users via multiple senses such as visual, auditory and tactile cues, and their combinations. The multisensory display modes reach a multitude of user profiles with varying degrees of accessibility and usability needs. Accordingly, mobility and navigation applications augment numerous users with new control and interaction capabilities for enhanced independence, quality of life, and social connectivity [21,31]. The widespread adoption and continuous improvement of ubiquitous navigation applications with novel interfaces are therefore considered to have a tremendous positive impact on building an inclusive society as a whole. New research and development in hardware, software and human-computer interactions (HCI) are hence critical to broaden the use cases and user profiles of inclusive mobility and navigation applications.

One of the limitations of current A-GPS based navigation applications is the reduced accuracy while fixating positional data. This might result in accuracy errors between 10 and 655 metres, depending on the landscape characteristics and GPS signal quality [62,63,106]. Consequently, even under the most satisfactory conditions where the A-GPS accuracy is within 10-metres of proximity, navigation applications cannot help the users pinpoint their final destination. In other words, the current applications are successful in taking the users towards an approximate proximity of the desired destination at macro-level. Nonetheless, they lack micro-level navigation capabilities to finalise the users' journeys. In the current research, this will be referred to as the last 10-metres problem. For example, Apple Maps completes a route by announcing that "the destination is on your left"; however, it cannot help the users to locate targets such as an entrance in the last 10-metres. Overall, this can significantly reduce

the mobility of some user profiles (e.g. visually impaired or elderly) while negatively affecting the quality of life across various users. Tackling this problem necessarily requires novel navigation systems, both in hardware and software, with a higher level of precision and seamless user interactions than the current A-GPS based navigation applications can provide. This raises new research challenges regarding how to develop and test interfaces for micro-level navigation (e.g. the last 10-metres problem) in HCI research.

For the sighted users, the last 10-metres problem can be solved rather intuitively because visual information is excessively available for guidance. Seeking visual cues in a complicated urban environment can, however, still be problematic in certain circumstances, such as driving a car [39]. If visual cues are restricted or absent to the user (e.g. a fireman rushing through heavy smoke [47], or a visually impaired user surveying his/her surroundings with a white cane), the last 10-metres problem becomes a great physical and mental challenge. The main barrier of current A-GPS navigation applications attempting to overcome the last 10-metres problem is that the physical to digital infrastructures are not yet capable of supporting positional accuracy with higher granularity. There are, however, recent instances of improvements in technological infrastructures, such as indoor tracking and navigation systems [32], which enable positional information with higher resolution. Technological developments are accelerating at a rapid rate. It will not be long until emerging technologies, for example those that utilise 5G, IoT devices and other sensors in smart buildings and cities, will be able to surpass the current limitations to positional accuracy at the micro-level [1,2,70,92,111,17,19,20,22,34,37,38,57]. Eventually, this will require novel ways of interacting with emerging navigation technologies.

For this, the current research investigated cross-modal displays with unisensory and multisensory modes, and studied their performance in micro-level navigation tasks. In a pilot study, a micro-level navigation task was set up in a classroom. Blindfolded sighted users were asked to find the door through randomly configured desks as obstacles. This created a scrambled path approximately 10-meters long from their

start position to the desired destination. A second navigation task was also arranged. Here, users had to find the door and then walk to the cafeteria. With low-fidelity prototypes, micro-level navigation cues were provided via non-speech positional auditory and tactile, or speech and non-positional tactile cues. This feedback indicated the direction users should move towards when they were getting further away from the target (e.g. becoming lost). Prior to navigation, users first received speech-based commands for macro-level navigation. For the first task, this was 'The route to the door is set. Turn right and in 10-metres, you will arrive in your final destination'. For the second task, this was 'The route to the cafeteria is set. Turn right and in 10-metres, you will arrive at the door. Then, turn left and in 100 metres, you will arrive in your final destination.' Interviews with seven users, who were tested with these low-fidelity prototypes, revealed that non-speech based positional auditory and tactile cues were more effective than speech and non-positional tactile cues in micro-level navigation. In contrast, speech and non-spatial tactile feedback were preferred for macro-level navigation. Users also reported that they preferred occasional feedback rather than constant stimulation.

Two cross-modal displays were prototyped. The prototypes were made of visual-to-auditory and visual-to-tactile sensory substitution devices (SSDs), which can deliver non-speech based positional sonifications and two-dimensional tactile cues. In a series of two experiments, the performances of sonification (unisensory), tactile (unisensory), kinaesthesia (unisensory), sonification-kinaesthesia (multisensory) and sonification-tactile (multisensory) display modes were studied in spatial tasks that relate to micro-level navigation. Experiment I examined Sonification-Kinaesthesia Prototype, which enables users to actively explore their surroundings from first-person perspective. Similarly, Experiment II evaluated Sonification-Tactile Prototype, which enables users to survey the external environment from a bird's-eye view. The performance of multisensory display modes of Sonification-Kinaesthesia and Sonification-Tactile Prototypes were compared to their unisensory modes in the same environment, where participants were asked to complete a spatial path integration task. This task included the completion of two distinct routes. These routes consisted of segments within 10-metres proximity of the participants. Their completion was

hence analogue to the last 10-metres problem. Participants' navigations were further tracked by motion trackers. The data from motion tracking was analysed for measuring the accuracy and precision of micro-level navigation performance. Participants were also interviewed about the navigation strategies they preferred with display modes. The findings indicated that cross-modal tactile cues were the most successful in delivering spatial information in micro-level navigation tasks.

4.2 Background

Navigation covers a wide range of spatiotemporal scales, such as wayfinding and locomotion [64,65]. Wayfinding is similar to macro-level navigation and refers to knowing how to get to a destination [64]. Locomotion is similar to micro-level navigation and refers to gaining distance estimates while navigating locally through obstacles [81,82]. The relation between micro- and macro- level navigation, hence the last 10-metres problem, is therefore relevant when considering how to create effective navigation technologies. The levels of navigation consist of egocentric (e.g. self-centred distance estimations) and allocentric (e.g. angular estimations between targets or landmarks) representations [64], which complement each other for robust spatial judgements [14]. This co-existence is achieved via spatial updating between the egocentric and allocentric representations [66,91]. Therefore, research targeting assisted navigation technologies should have a holistic view of navigation to tackle the 10-metres problem. In making the prototypes and designing the experimental procedure, we considered (i) why cross-modal displays would be feasible mediums to convey cues for micro-level navigation, and (ii) how multisensory combination of spatial cues could minimise the disadvantages of unisensory displays (e.g. sensory adaptation) and multisensory display modes (e.g. switching cost).

4.2.1 Cross-Modal Displays

Sensory substitution devices (SSDs) are considered cross-modal displays that are built with the principles of how complementary cross-modal cues correspond to each other [48]. In this way, sensory substitution techniques make it possible to represent or

compliment the same sensory information with different sensory channels. Users, for example, can acquire visual information by means of non-speech sonifications [61] or two-dimensional tactile cues [9], and auditory information by means of vibrotactile cues [15,23,26,73]. Sensory substitution techniques also enable interaction with novel forms of information by transforming, extending and augmenting our perceptual capacities [8,55,56]. For this reason, SSDs are classified as sensory augmentation devices, whereby users could augment their sensory abilities with additional inputs such as thermal imaging [69]. Furthermore, as they are compatible with flexible sensory input, sensory substitution techniques provide valuable tools for researchers to examine cross-modal displays in a variety of use cases such as intuitive user interactions [58] and assistive technologies [55,59,86]. In the current research, we assert that the spatial cues from future micro-level navigation applications could be displayed via sensory substitution techniques and hence studied with cross-modal displays.

4.2.2 Multisensory Combination

Our overall multisensory experience enables navigation in the physical environments with a combination of intact senses [16,30,49,54,68,100]. Despite the advancements in understanding multisensory processing and its evolutionary benefits in human cognition, multisensory display modes are studied in less detail than the unisensory modes in HCI research [75,76,94]. It is argued that unisensory modes outperform their multisensory alternatives because of the switching cost between different sensory sources [52]. That is, sensitivity to the sensory information and the reaction times to act are reduced when users switch their attention between distinct information channels (e.g. visual, auditory and tactile cues) that make of the multisensory display mode [50,83,93]. On the other hand, research also shows that continuous sensory information from the same sensory channel via unisensory displays lead to sensory adaptation, which reduces overall sensitivity in change detection [3,33,35]. This further suggests that multisensory displays could indeed be beneficial for navigation applications as users would be more alert to different sensory cues instead of suffering from the consequences of sensory adaptation.

In exemplars of cross-modal displays with multisensory modes, complementary cues are utilised to decrease the cognitive load of the users [40–42]. It is suggested that such displays could minimise the effects of sensory adaptation and optimise the switching cost, thereby enhancing overall user experience. This is achieved via a multisensory process called multisensory combination. Essentially different than multisensory integration (for a detailed review, see [87]), multisensory combination processes complementary cross-modal cues to form an improved multisensory judgement than unisensory processing [13,27,99]. For example, multisensory combination allows an object's shape to be consciously estimated via visual inspection and also touching the areas that cannot be seen [71]. Accordingly, the current research investigated whether cross-modal spatial information could be combined efficiently to minimise sensory adaptation and optimise switching cost. Utilising multisensory combination, the cross-modal display prototypes could provide rich egocentric and allocentric cues, which would enhance overall navigation performance.

4.3 Experimental Investigation

One navigational process, which investigates locomotion and wayfinding together, is path integration. Path integration is a fundamental mechanism of spatial navigation that enables navigators to memorise paths in segments (for a detailed review, see [29]). Path integration creates mental representations (i.e. egocentric and allocentric) via spatial updating during locomotion [28,109]. In this way, navigators can subsequently update their position with respect to definite landmarks (e.g. the start of a path segment) in wayfinding [28,109]. These further suggest that micro- and macro- level (i.e. locomotion and wayfinding respectively) navigation applications should be conceptualised as a whole rather than independent modules in tackling the last 10-metres problem. For example, it was evidenced that visual and interoceptive (e.g. kinaesthesia) cues, which provide allocentric and egocentric spatial representations respectively [80], combine for path integration [100]. Similarly, it was found that auditory and tactile sensory substitution techniques provide egocentric and allocentric cues respectively [77]. This supports the idea that cross-modal displays

with unisensory and multisensory modes can be studied with path integration to address the last 10-metres problem. Moreover, it is further suggested that multisensory combination can be operationalised via cross-modal displays so that both egocentric and allocentric representations could be combined via sensory substitution techniques for enhanced navigation.

Path integration is often studied with a triangle completion task [100,109]. In the exploration phase, participants survey three targets that form the three corners of a triangle. In the task phase, without any environmental cues, blindfolded participants are then asked to walk through a combination of the legs of the triangle (e.g. walk to the third target or walk to the third target via the first target). Varying the leg combinations for the task phase allows studying both egocentric (i.e. when the target can be reached directly) and allocentric (i.e. when the target can only be reached via passing through another target) routes. Accuracy and precision can then be calculated with respect to the distance between the target and the end point of the participants.

For the series of two experiments presented in the current research, a triangle completion task of one egocentric and one allocentric route was deployed to investigate the navigation performance of the cross-modal display prototypes in unisensory and multisensory modes. The exact same experimental procedures were followed in both experiments. Three targets were placed in a triangular configuration in a 10x7m laboratory so that navigating the routes in path integration would be analogous to that of the last 10-metres problem. The laboratory was equipped with a motion tracking system to record participants' navigations during the task phase of the triangle completion task. The experimental procedure allowed the investigation of a combination of cross-modal cues (sonification, kinaesthesia and tactile) in two display modes (unisensory and multisensory), spatial representations and updating (egocentric and allocentric) and viewing perspectives (first-person and bird's-eye). While each experiment was run separately as within-subject design, a between-subject comparison between the two studies was also made for future comparisons.

4.3.1 Experimental Hypotheses

It was hypothesised that the use of cross-modal displays in a multisensory mode would minimise sensory adaptation and optimise the switching cost, thereby increasing overall sensitivity and change detection (H1). That is, accuracy (H1.1) and precision (H1.2) of navigation in the multisensory mode would be higher than unisensory modes in both cross-modal prototypes. Furthermore, it was hypothesised that conveying complementary cross-modal cues would benefit from multisensory combination (H2.1). This would consequently formulate the basis of enhanced performance predicted in H1. Following this, it was hypothesised that user strategies would reflect the benefits of multisensory combination (H2.2).

4.4 Methods

4.4.1 Participants

In total, 24 right-handed participants were recruited for Experiment I (12M, mean age of 24.75, SD = 5.7), and 30 right-handed participants were recruited for Experiment II (15M, mean age of 22, SD = 3.2) from the University of Bath, UK. Participants were screened for normal vision, audition and touch sensations. They were novice to the cross-modal displays and prototypes used in the experiments. Participants were only allowed to take part in one of the experiments. None of the participants had been to the laboratory where the experiments took place previously. This ensured that they would not have any spatial biases (e.g. size and shape of the laboratory). Each experiment took approximately 3.5 hours to complete with small breaks between experimental conditions. Participants were informed that they would be blindfolded during the experiment with the exception of breaks, which were held outside the laboratory. Prior to the onset of the experiments, participants provided informed consent and they were also debriefed at the end. Participants were reimbursed £5 for their time. The experiments were approved by the University of Bath Psychology Department Ethics Committee (Experiment I: 16-180 and Experiment II: 17-273).

4.4.2 Apparatus

4.4.2.1 Cross-Modal Display Prototypes

Two cross-modal displays with unisensory and multisensory modes were prototyped. In making these prototypes, an auditory-to-visual sensory substitution device, namely The vOICe [61], was utilised for creating non-speech based sonifications. The cross-modal conversion The vOICe algorithm is based on the relations between elevation and pitch, brightness and loudness, and stereoscopic and horizontal positioning. A tactile-to-visual sensory substitution device, namely BrainPort [108], was also utilised for conveying two-dimensional tactile cues. The cross-modal conversion BrainPort algorithm creates patterns displayed on an electrotactile intra-oral display, where the intensity of individual electrodes represents brightness. As The vOICe and BrainPort are commercially available, we followed very simple procedures while making the prototypes for ease of replication.

Sonification-Kinaesthesia Prototype

The Sonification-Kinaesthesia Prototype utilises The vOICe algorithm as a cross-modal display. It enables users to actively explore an external environment via sonifications with or without kinaesthetic cues from a first-person perspective. The prototype consists of an adjustable helmet with reflectors for motion tracking, a blindfold necessary for the triangle completion task, and neckband stereo headphones as an output device to listen to the sonifications (Figure 4.1). It was previously evidenced that head-mounted cameras lead to higher navigation performance than hand-held cameras with sonifications [12]. The prototype therefore consists of a USB camera (ELP 480P with 120° view) for sensory input, which is attached precisely above the middle of the blindfold. The prototype is connected to a pocket computer for ease of carriage so that users are able to walk freely without any wired connections. The vOICe is run at default settings (1s scan rate, normal contrast, foveal view off) via the pocket computer to sonify the real-time camera feed. Sonification-Kinaesthesia Prototype was evaluated in Experiment I.

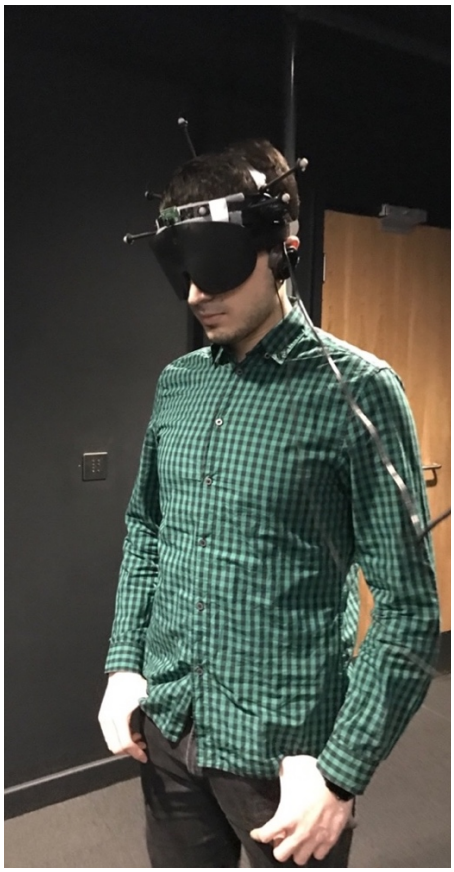


Figure 4.1 displays a user wearing Sonification-Kinaesthesia Prototype.

Sonification-Tactile Prototype

The Sonification-Tactile Prototype utilises The vOICe and BrainPort as auditory and tactile cross-modal displays respectively. It enables users to explore an external environment via sonifications, two-dimensional electro-tactile cues or their multisensory combination from a bird's-eye view. The prototype consists of an enclosed body (40x40x40cm) to control lighting and other environmental factors, an adjustable scaffolding mechanism to attach BrainPort's camera, a PC to run BrainPort and The vOICe simultaneously, an adjustable helmet with reflectors for motion tracking, a blindfold necessary for the triangle completion task, and neckband stereo headphones as an output device to listen to the sonifications (Figure 4.2). Users receive electro-tactile cues from BrainPort's intra-oral display. Inside the enclosed body, there is an A5 sized viewing platform (14.8x21 cm), where the bird's-eye view maps could be placed 23cm away from the camera (see Bird's-eye View Maps for more details).

Unlike The vOICe, which is purely a software programme and compatible with most camera connected devices, BrainPort is intact. That is, it consists of its own helmet, camera, processor and intra-oral display, and therefore cannot be connected to an external camera. This technically prevented the Sonification-Tactile Prototype to convey the same visual input in the multisensory mode from the first-person perspective of one camera. Even if two cameras (one for each cross-modal display) were to be placed adjacently and used simultaneously, they would misalign the cross-modal cues spatially and, to a degree, temporally. As this misalignment would lead to multisensory response depression [97,98], using two separate cameras was not considered as an option. Previous research shows that sonifications and tactile maps from an aerial view could convey spatial cues for successful navigation [77]. Similarly then, the Sonification-Tactile Prototype was designed to deliver pre-recorded sonifications and real-time tactile cues of matching camera feeds in unisensory and multisensory modes.

The Sonification-Tactile Prototype assisted the consistent delivery of sonification and tactile equivalents of the identical bird's-eye map in Experiment II. The maps were sonified and recorded using The vOICe algorithm at default settings (1s scan rate, normal contrast, foveal view off) prior to the experiment. Unlike The vOICe, BrainPort cannot be uploaded with pre-recorded stimulus. The live feed from its camera was used to display the maps via electro-tactile cues during the experiment at zoom 33°, invert off, contrast high, lighting low, tilt 25° settings. In this way, it was consistently guaranteed that BrainPort's camera view matched the previously sonified image for successful multisensory combination. This was further ensured via BrainPort's HTML based interface.

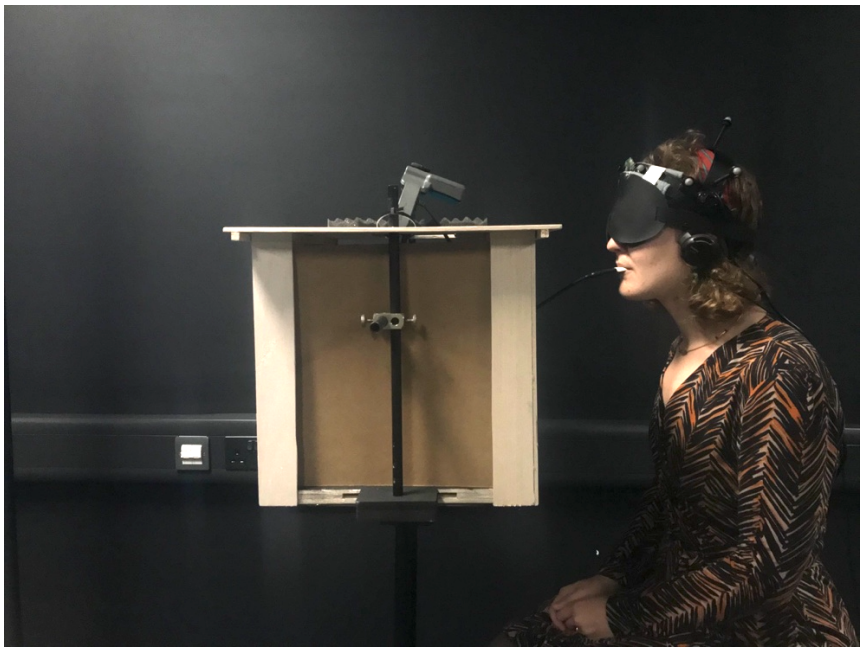


Figure 4.2 displays a user using the Sonification-Tactile Prototype in the multisensory mode.

4.4.2.2 Motion Tracking

The laboratory where the experiments took place was equipped with Vicon Bonita motion tracking system with 8 infrared cameras. The motion trackers tracked 5 reflectors that were placed non-linearly on the adjustable helmet (Figure 4.1, Figure 4.2). The Vicon system was controlled through a custom-made script in Python 3.0 utilising Vizard tracking libraries.

4.4.3 Stimulus Design

For the triangle angle completion task, three targets of equal size with 60 cm height and 50x50cm base were crafted (Figure 4.3). These targets, a pyramid, a cylinder and a rectangular prism, were configured in a triangular formation (Figure 4.4). The targets were made of white cardboard, which created a strong contrast on the black coloured floor and walls of the laboratory, for ease of detection with the cross-modal display prototypes. An alternative triangular configuration was used for training so that participants would not be biased to the experimental configuration of the targets (see Figure 4.5 for the training configuration of the targets from the bird's-eye view).

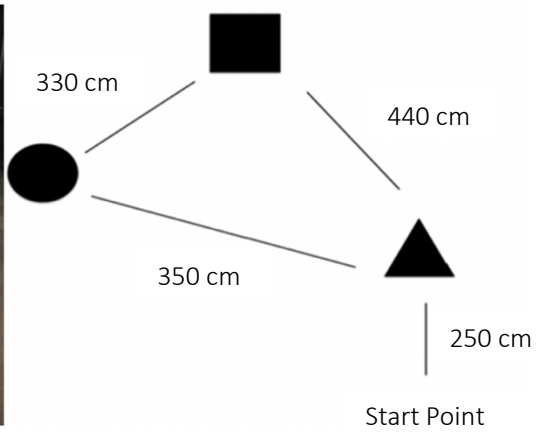


Figure 4.3 (left) and **Figure 4.4** (right) display the three targets and their triangular configuration used in the experiments. The start point indicates the point where participants started exploring the targets (exploration phase) and also started the navigation tasks (task phase). Blue-taped areas on the floor show the training configuration of the targets while the start point is invisible.

4.4.3.1 Bird's-eye View Maps

While Experiment I and Experiment II used the same target configurations for training and experimentation, in Experiment II, these configurations were transposed to a bird's-eye view map. The positions of the start point and the three targets were digitally recreated to scale using AutoCad [6], and indicated by a white square on a black background (Figure 4.5).

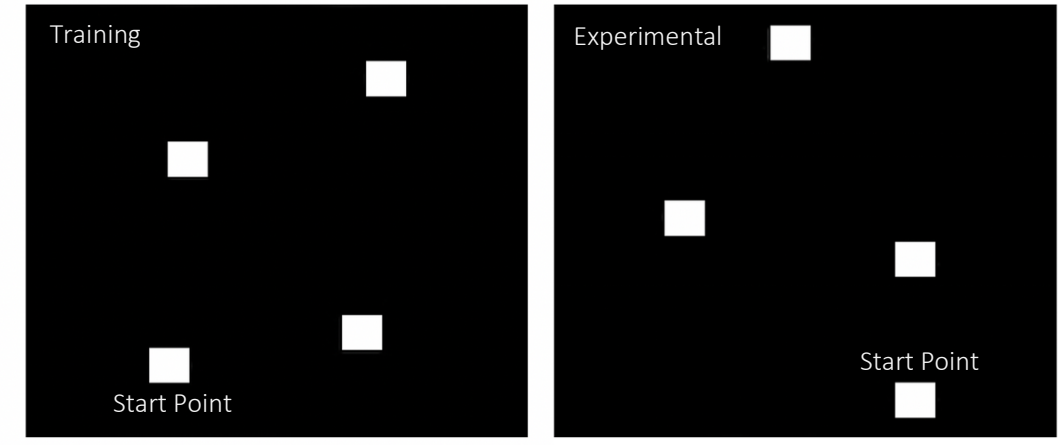


Figure 4.5 illustrates the birds-view maps of training configuration (left) and experimental configuration (right). The start points are respectively on the bottom left and the right corners.

4.4.4 Triangle Completion Task

The triangle completion task consisted of two routes (Figure 4.6). The egocentric route required participants to walk directly to an end target (i.e. cylinder) and the allocentric

route required participants to walk to an end target (i.e. cylinder) by passing through another target (i.e. pyramid). In this way, participants were never guided in these routes by the experimenter during the exploration phase. Participants were always oriented towards the first target they would navigate towards. A secondary experimenter removed the targets quietly prior to the task phase and participants were only stopped if they were to collapse into walls.

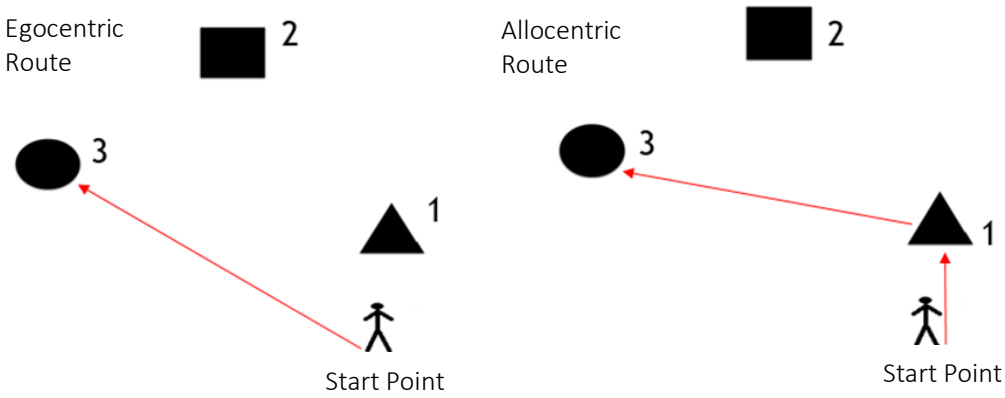


Figure 4.6 illustrates the egocentric route (left) and allocentric route (right) of the triangle completion task.

4.4.5 Experimental Conditions

Experiment I and Experiment II followed the exact same procedure with the exception that the prototypes participants used were different. In both studies, there were three experimental conditions (two unisensory and one multisensory) depending on the display mode used to survey the environment (Table 4.1). The three conditions were counter-balanced in six groups across participants with approximately equal number of males and females in each group.

Table 4.1 Table to show the procedures used to present each display mode in all phases of the first and second experiment

	Condition	Exploration phase	Task phase
Experiment 1	Sonification mode	Participants stands still at the start point and explores the three targets.	All objects are removed and display modes are deactivated. Participants are asked to navigate to Target 3 (egocentric) or Target 1 and then Target 3 (allocentric). Their motion is tracked. Motion tracking begins when they start navigating and ends when participants announce that their arrival at Target 3. They are then returned back to the start point via a scrambled path.*
	Kinaesthesia mode	Participants are guided to Target 1, Target 2 and Target 3 in this order. They are returned back to the start point via a scrambled path.	*
	Sonification-Kinaesthesia mode	Participants are always guided to Target 1, Target 2 and Target 3 in this order. In this mode, participants are also allowed to listen to the soundscapes. They can look around and gain environmental cues via the sonifications. They are returned back to the start point via a scrambled path.	*
Experiment 2	Sonification mode	Participants listen to the pre-recorded sonifications of the aerial map.	*
	Tactile mode	Participants receives the tactile feedback of the aerial map on their tongues.	*
	Sonification-Tactile mode	Participants receive both the sonification and tactile feedback simultaneously.	*

*During the task phase the same procedure is repeated across every display more for both experiments.

4.4.5.1 Experiment I: Tested Kinaesthesia, Sonification and Sonification-Kinaesthesia Display Modes from First-Person Perspective

In Experiment I, the three conditions were sonification (unisensory), kinaesthesia (unisensory) and sonification-kinaesthesia (multisensory). In the sonification condition, participants stood at the start point and actively explored the environment. In the kinaesthesia condition, the prototype was muted, and participants were guided to each target (in the order of 1-2-3, see Figure 4.6) from the start point. The experimenter led them back to the start point via a scrambled path. In the sonification-kinaesthesia condition, participants were guided to each object in the same order as the kinaesthesia condition while being able to freely explore the environment via sonifications in real-time. They were returned back to the start point via a scrambled path.

4.4.5.2 Experiment II: Tested Tactile, Sonification and Sonification-Tactile Display Modes from Bird's-Eye View Perspective

In Experiment II, the three conditions were sonification (unisensory), tactile (unisensory) and sonification-tactile (multisensory). Participants stood by the Sonification-Tactile Prototype to either listen to the sonification of pre-recorded bird's-eye view map, feel the electro-tactile map or used both of the display modes simultaneously. Surveying the map lasted for 10 seconds in every condition and the name of the targets were announced to the participants from left to right prior to the initiation of sonification and/or tactile display mode.

4.4.6 Experimental Procedure

The experimental procedure consisted of two phases: training and navigation experiment. Participants were initially welcomed in a different room than where the navigation experiment would take place. They provided their consent and were then briefed about the experiment. They were allowed a short break between phases of training and experimental conditions.

4.4.6.1 Training

Participants were given an online presentation about the main principles of cross-modal displays they would use during the experiment. This presentation was then followed by a quiz of 10 sonifications for every participant. In Experiment II, an additional quiz was completed with tactile display mode. In these quizzes, blindfolded participants were asked to recognise simple shapes, such as a circle, and the orientation of lines and dots with four-alternative force choices from the sonifications and tactile feedback. After each question, they were given a brief feedback to reinforce their understanding of sensory substitution and how the cross-modal displays worked. Participants were expected to have at least 80% success rate before proceeding with the rest of the experiment. After the successful completion of quizzes, participants were introduced to the prototypes they would use and how they worked. They were then taken to the laboratory where the navigation experiment would take place. This procedure approximately took 30 minutes.

The second phase of training happened in the laboratory and aimed to equip blindfolded participants with hands-on practice with the prototype they would use. In order to familiarise them with the triangle completion task, participants practiced a pair of egocentric and allocentric navigation tasks with each display mode in the training configuration. Further feedback was given to the participants and their navigation was corrected. This feedback additionally established the distance calibrations between physical targets and their equivalent representations conveyed via the prototypes. Participants' navigation was not recorded with motion trackers during training. This procedure approximately took 45 minutes.

4.4.6.2 Navigation Experiment

In total, participants completed 10 pairs (i.e. one egocentric and allocentric route) of triangle completion tasks with each display mode. Every task pair started with participants exploring the target configuration with the given display mode of the condition (exploration phase). Then, the prototype was paused to prevent participants receiving feedback during the task phase. Motion tracking started as soon

as participants started navigating and ended when participants announced that they completed the route. The order of routes was altered after each trial pair. In total, each participant completed 60 trials of triangle completion tasks. The navigation experiment took approximately 2 hours. At the end, participants were taken outside the laboratory to complete a questionnaire on the strategies they used for each display mode.

4.5 Results

Performance from the three display modes was examined with respect to constant and variable errors, which refer to navigation accuracy and precision respectively. Constant error represents a systematic navigational bias with respect to the available spatial information [18]. Variable error measures how robust the perceptual judgements are as a result of unisensory processing or multisensory combination [18]. Accordingly, constant error is expected to be reduced in enhanced navigation trials, and variable error should be lower when the spatial representations are robust. Six pairs of constant and variable error values were calculated for each participant with respect to the display mode and route type.

To calculate these errors, the 3D coordinates obtained from motion-tracking data were processed using MATLAB [60] and Psychtoolbox command library [11]. A bivariate normal distribution was fitted to the finishing coordinates of each navigation task. This enabled the estimation of the mean and variance of x and y coordinates of where navigations were finished. FASTCMD algorithm [88] from MATLAB Libra toolbox [105] was used for a robust estimation of these values, with the assumption of 1% outlier values where α =.99. Constant error was thereby calculated as the distance between the centre of the fitted bivariate distribution and the correct position of the target. Similarly, variable error was calculated as the sum of the variance of x and y coordinates of the fitted bivariate distribution. The analyses of these errors will be reported separately for each study. The quantitative data analysis was completed with SPSS25 [46] and the qualitative analysis was carried out with two coders using ATLAS.ti

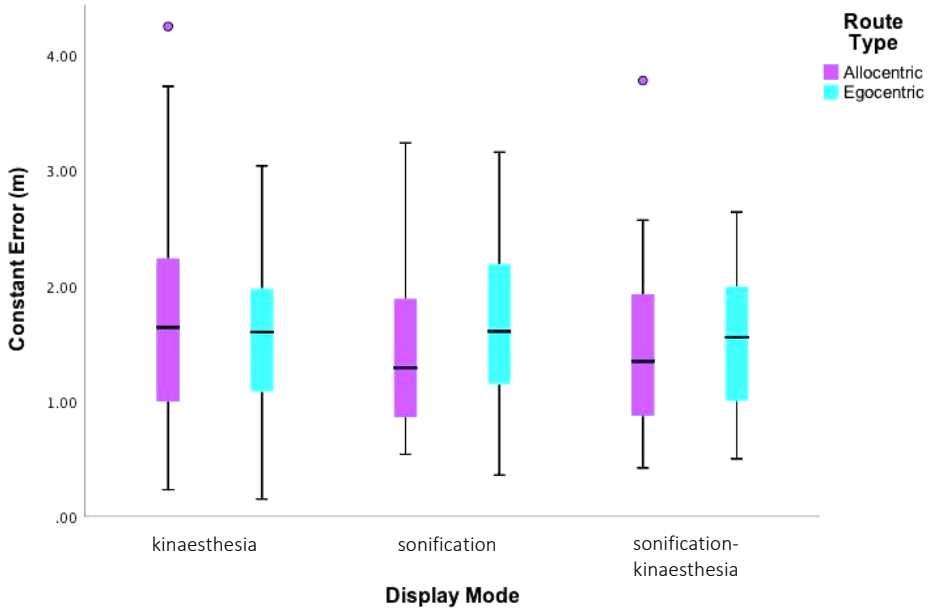
[5]. Standard deviations are indicated with ± throughout the paper unless otherwise was indicated.

4.5.1 Experiment I: Tested Kinaesthesia, Sonification and Sonification-Kinaesthesia Display Modes from First-Person Perspective

4.5.1.1 Constant Error

The average constant errors for kinaesthesia, sonification and sonification-kinaesthesia modes were 1.6m ± 0.7 (egocentric), 1.8m ± 1.0 (allocentric), 1.6m ± 0.8 (egocentric), 1.5m ± 0.8 (allocentric), 1.5m ± 0.6 (egocentric), and 1.5m ± 0.8 (allocentric) respectively (Boxplot 4.1).

A repeated measures ANOVA was carried out to compare constant error values between the interaction of display mode (sonification, kinaesthesia and sonification-kinaesthesia) and route type (egocentric and allocentric). The analysis showed no significant main effect of display mode, F(2,46) = 0.752, *p = .477, partial* $\eta^2 = 0.032$; route type, F(1,23) = 0.15, *p = .904, partial* $\eta^2 = 0.001$; and no significant two-way interaction between display mode and route type, F(2,46) = 1.339, *p = .272, partial* $\eta^2 = 0.055$.

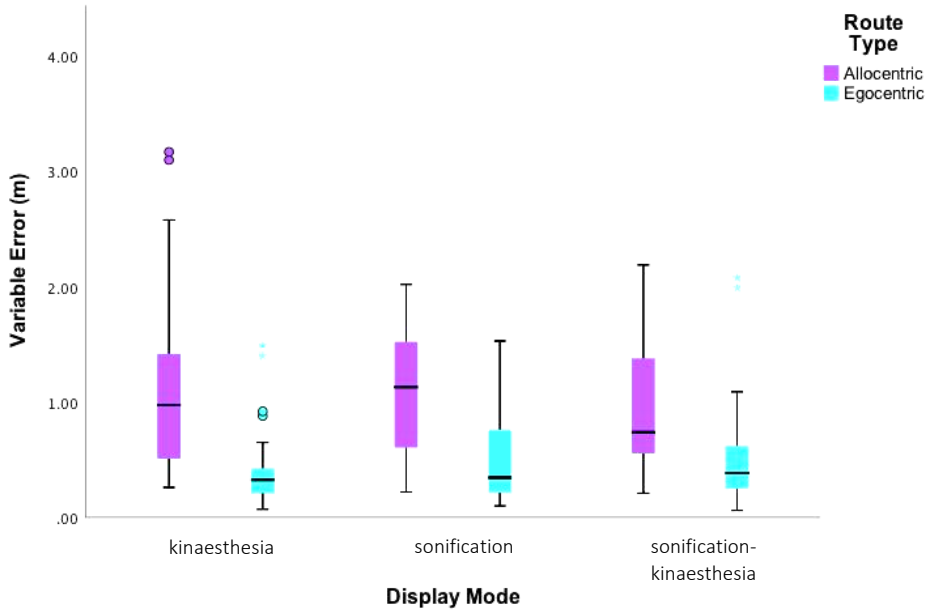


Boxplot 4.1 shows constant errors in metres for kinaesthesia, sonification and sonification-kinaesthesia display modes with respect to egocentric (cyan) and allocentric (magenta) routes.

4.5.1.2 Variable Error

The average variable errors for kinaesthesia, sonification and sonification-kinaesthesia modes were 0.4m ± 0.4 (egocentric), 1.2m ± 0.8 (allocentric), 0.5m ± 0.4 (egocentric), 1.4m ± 1.5 (allocentric), 0.5m ± 0.5 (egocentric), and 1.0m ± 0.6 (allocentric) respectively (Boxplot 4.2).

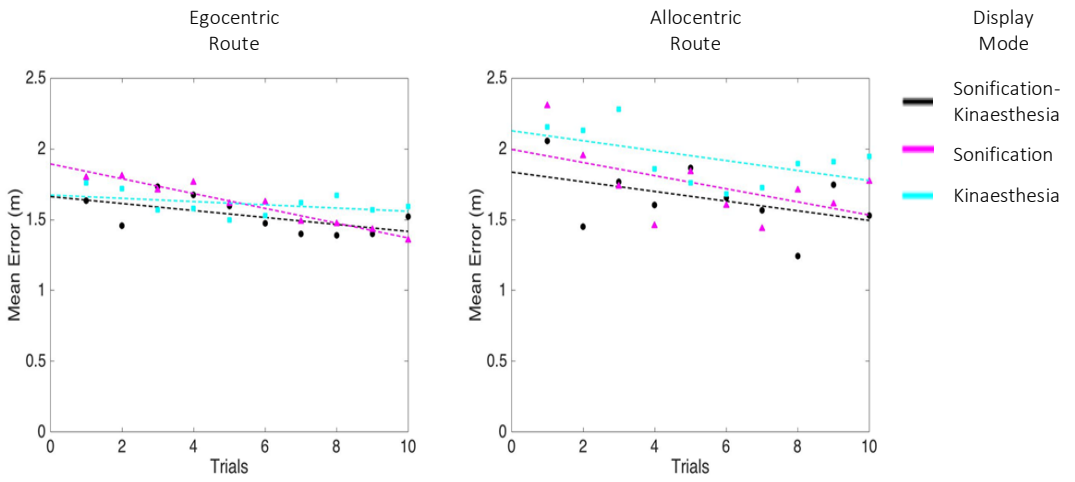
A repeated measures ANOVA indicated no main effect of display mode $F(2,46) = 0.680$, $p = .507$, *partial* $\eta^2 = 0.029$, and no significant two-way interaction between the display mode and route type, $F(1.448,33.310) = 0.963$, $p = .389$, *partial* $\eta^2 = 0.040$. However, a significant main effect of route type was found, $F(1,23) = 34.846$, $p<.001$, $\eta^2 = 0.602$. That is, the variable error from the allocentric route (M = 1.2m ± 0.2) was significantly higher than the egocentric route (M = 0.5m ± 0.1) with a mean difference of 0.7m (95% CI [0.438, 0.912], $p <.001$). Further Bonferroni corrected pairwise comparisons revealed that, while the variable error from all display modes was on average higher in the allocentric route, this difference only reached a significant level in the kinaesthesia mode with a mean difference of 0.8m (95% CI [0.253, 1.209], $p = .001$).



Boxplot 4.2 shows variable errors in metres for kinaesthesia, sonification and sonification-kinaesthesia display modes with respect to egocentric (cyan) and allocentric (magenta) routes.

4.5.1.3 The Effect of Learning

Even though participants were not given any feedback during the triangle completion tasks in Experiment I, an improvement in their navigation abilities with Sonification-Kinaesthesia Prototype was observed. To test whether this reached any significance level, a directional linear regression for egocentric and allocentric routes was calculated separately over the 10 trials completed with each display mode. This analysis examined whether the average constant error decreased as a function of the number of trials (Graph 4.1).



Graph 4.1 represents the effect of learning across 10 trials for egocentric route (left) and allocentric route (right). The display modes are indicated by black (sonification-kinaesthesia), magenta (sonification), and cyan (kinaesthesia).

Sonification mode showed a significant improvement in both egocentric, $F(1,9) = 132.415$, $p < .001$, and allocentric $F(1,9) = 3.487$, $p = .049$, routes. Sonification-kinaesthesia mode also yielded a significant improvement in the egocentric route, $F(1,9) = 4.433$, $p = .034$. The effect of learning was not observed with sonification-kinaesthesia mode in the allocentric route, $F(1,9) = 2.074$, $p = .094$, and with kinaesthesia mode in the egocentric route, $F(1,9) = 1.672$, $p = .132$ and in the allocentric route, $F(1,9) = 3.242$, $p = .54$.

4.5.1.4 Qualitative Strategies

In the kinaesthesia mode, 92% of the participants reported that they counted their steps while being guided by the experimenter. This helped them to roughly estimate the distance between Target 1 and Target 3. Participants did not like 'the lack of feedback' because they could not know they were at the target. Lack of feedback also led to 'a sense of dependency'. Eighty-seven percent of participants also mentioned that estimating distances with kinaesthesia was easier than understanding the angles between the targets.

Eighty-two percent of participants described the sonification mode as a way of 'visually imagining' a 'mental map of where the targets were with respect to [my] location and with respect to each other'. They used the pitch and volume of sonifications to 'figure out the angles' between targets. Overall, 90% of them reported to be 'more confident on the angle'. They also reported that they actively moved their heads right-left and up-down to understand the relations between targets. Even though the conditions of Experiment I were counterbalanced, 76% of participants additionally reported that they 'tried to remember step counts from previous tasks' for distance estimations.

Eighty-four percent of participants resembled the sonification-kinaesthesia mode to 'looking around with sensory substitution while walking'. Sixty-two percent of participants found the sonification-kinaesthesia mode the easiest as it provided 'a combined picture' with both the 'angle and distance feedback' to 'update [my] knowledge of the location of targets'. This was particularly mentioned for 'figuring out the angle between Target 1 and Target 3'. Understanding how Target 3 was placed in relation to Target 1 made participants correct the angles they took in the allocentric task. Being able to view both Target 2 and Target 3 from Target 1 via sonifications also helped them 'work out the distance from Target 1 to Target 3'. The Sonification-kinaesthesia mode was also found to be the most secure and reliable by 67% of participants. One participant particularly mentioned that the 'information was

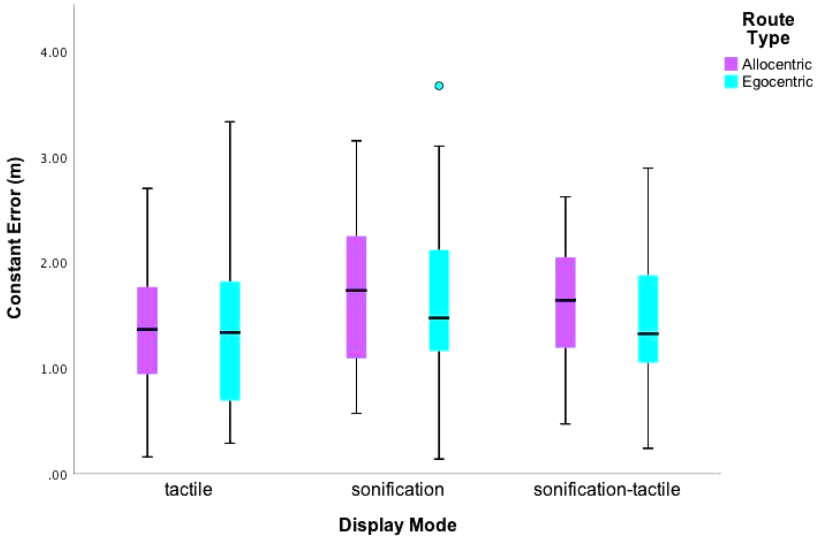
overwhelming', and another specifically wrote that they were 'switching between just counting [my] steps and using The vOICe'.

Overall, when asked to rank the display modes in ease of use for the navigation task, 62% of participants found sonification-kinaesthesia mode the easiest, 54% found sonifications the second easiest, and 67% found kinaesthesia mode the hardest.

4.5.2 Experiment II: Tested Tactile, Sonification and Sonification-Tactile Display Modes from Bird's-Eye View Perspective

4.5.2.1 Constant Error

The average constant error values for tactile, sonification and sonification-tactile modes were 1.4m ± 0.8 (egocentric), 1.4m ± 0.7 (allocentric), 1.6m ± 0.8 (egocentric), 1.7m ± 0.8 (allocentric), 1.4m ± 0.7 (egocentric), and 1.6m ± 0.6 (allocentric) respectively (Boxplot 4.3).



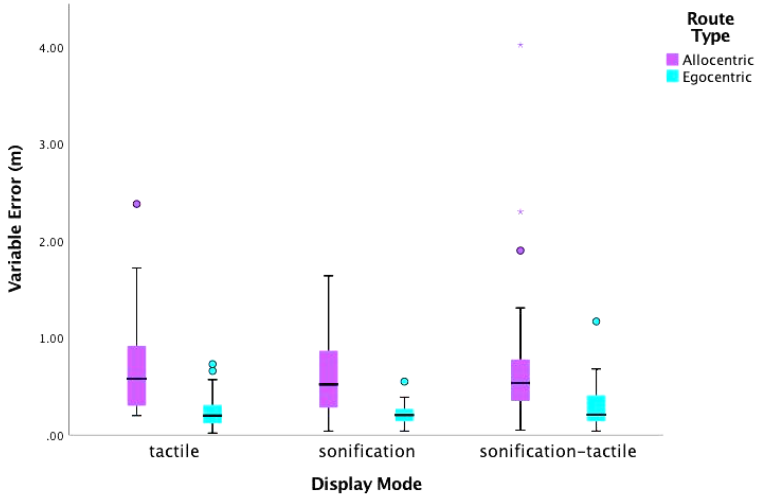
Boxplot 4.3 shows constant errors in metres for tactile, sonification and sonification-tactile display modes with respect to egocentric (cyan) and allocentric (magenta) routes.

A repeated measures ANOVA was carried out to compare constant error between the interaction of display mode (tactile, sonification and sonification-tactile) and route type (egocentric and allocentric). The analysis showed no significant main effect of route, $F(1,29) = 1.790$, $p = .191$, *partial* $\eta^2 = 0.058$, and no significant two-way

interaction between display mode and route type, F(2,58) = 1.237, *p = .298, partial* η^2 *= 0.041*. However, a significant main effect of display mode was found, F(2,58) = 5.861, *p = .005, partial* η^2 *= 0.168*. Further Bonferroni corrected pairwise comparisons were carried out to investigate this. Tactile mode (1.4m ± 0.1) led to significantly lower constant error than the sonification mode (1.6m ± 0.1) with a mean difference of 0.25m (95% CI [0.068, 0.427], *p = .005*). No other significant difference was found between the display modes.

4.5.2.2 Variable Error

The average variable errors for tactile, sonification and sonification-tactile modes were 0.2m ± 0.2 (egocentric), 0.7 ± 0.5 (allocentric), 0.2m ± 0.1 (egocentric), 0.6m ± 0.4 (allocentric), 0.3m ± 0.2 (egocentric), and 0.8m ± 0.8 (allocentric) respectively (Boxplot 4.4).



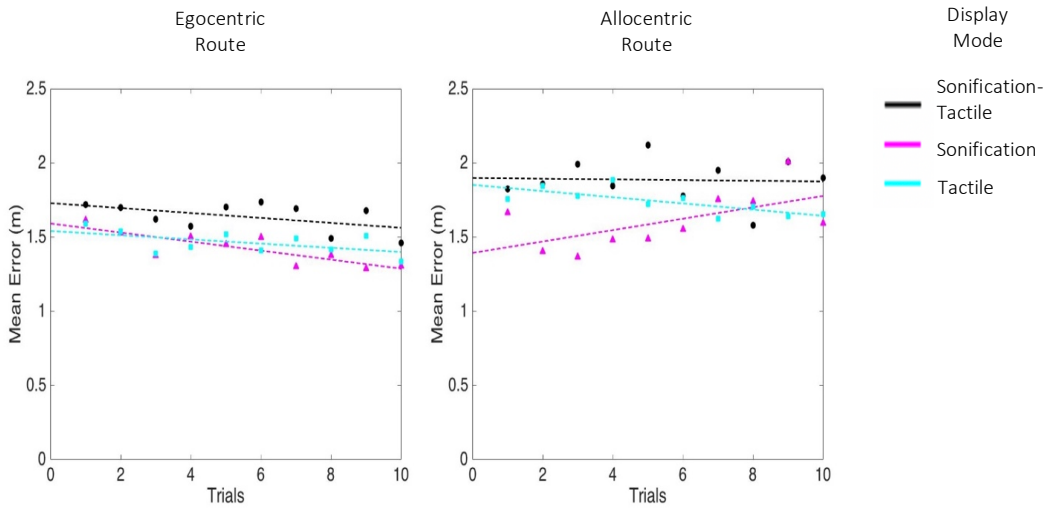
Boxplot 4.4 shows variable errors in metres for tactile, sonification and sonification-tactile display modes with respect to egocentric (cyan) and allocentric (magenta) routes.

A repeated measures ANOVA showed no significant main effect of display mode, F(1.512,43.838) = 1.271, *p = .283, partial* η^2 *= 0.042*, and no significant two-way interaction between display mode and route type, F(1.643,47.634) = 0.181, *p = .792, partial* η^2 *= 0.006*. However, a significant main effect of route type was found, F(1,29) = 38.915, *p < .001, partial* η^2 *= 0.573*. Further Bonferroni corrected pairwise

comparisons indicated that, in every display mode, egocentric route (0.3m ± 0.02) yielded significantly lower variable error than the allocentric route (0.7m ± 0.1) with an overall mean difference of 0.4m (95% CI [0.296, 0.584], *p < .001*).

4.5.2.3 The Effect of Learning

Similar to Experiment I, participants' navigation abilities with Sonification-Tactile Prototype were improved over trials. To test whether this reached any significance level, a directional linear regression for egocentric and allocentric routes was calculated separately over the 10 trials completed with each display mode. This analysis examined whether the average constant error decreased as a function of the number of trials (Graph 4.2).



Graph 4.2 represents the effect of learning across 10 trials for egocentric route (left) and allocentric route (right). The display modes are indicated by black (sonification-tactile), magenta (sonification) and cyan (tactile).

Participants showed a significant improvement in their navigation performance with the sonification mode for the egocentric route, F(1,9) = 17,217, *p = .003.* This was not significant in the allocentric route, F(1,9) = 4.511, *p = .066.* Participants showed a significant improvement with the tactile mode for the allocentric route, F(1,9) = 9.391, *p = .015*. This did not reach a significant level in the egocentric route, F(1,9) = 3.384, *p = .103.* Participants showed no significant improvements with the sonification-tactile mode for egocentric, F(1,9) = 2.878, *p = .128*, and allocentric, F(1,9) = 0.017, *p = .900*.

4.5.2.4 Qualitative Strategies

In the tactile mode, 94% of participants found it easier to 'judge relative locations' because this helped them 'mentally visualise' the sensory information and estimate the 'angles between targets'. Fifty-eight percent of participants especially noted that they actively moved their tongues over the tactile display mode up/down and left/right to figure out distance and angle cues respectively. Eight-five percent of participants also reported that they locally explored targets first and then created a global image. For example, they tried to 'quickly judge the approximate locations of each target and then go back to slowly work out more precise relative distances and angles'.

Seventy-three percent of participants found the sonification mode 'quite difficult to use'. Vertical alignments were particularly difficult to recognise whereas 'the horizontals involved less guess work'. The pitch and temporal differences in sonifications were dominantly used for distance estimations between targets. On the contrary, finding out the angles were 'a guess work'. One participant reported that sonifications resembled musical pitch, 'so the start point felt like a C, Target 1 a D and Target 3 an F. Therefore, I estimated that Target 3 must be twice as far to Target 1 than the start'. Similarly, another participant wrote that, 'I imagined a chromatic scale being proportional to the vertical distances between points and estimated that the time intervals between sounds were roughly equal'.

In sonification-tactile mode, 93% of participants clearly used the tactile mode for estimating the angles between targets and the sonification mode for distances to 'confirm [my] knowledge of the target layout'. They found 'using both devices simultaneously a bit difficult' and tended to focus one at a time, switching between the two frequently'. However, there were no particular order of using one display mode after the other. Eight-six percent of participants also reported that they used sonifications to 'mentally map a configuration' and 'visualise the location of targets more easily' in the tactile mode. This also provided them with 'horizontal and vertical resolution' in sonification and tactile modes. Overall, 17% participants reported that

they relied more on the tactile mode than sonifications because it gave a 'sense of space' and 'a more mapped out way of visualising the points', which required 'only a scaling factor'. A minority (1% of the participants) found sonifications more useful than the tactile mode.

Overall, when asked to rank the display modes in ease of use in navigation tasks, 53% of the participants found sonification-tactile mode the easiest, 47% found the tactile mode the second easiest and 73% rated sonifications as the hardest.

4.5.3 Between-Subject Comparison between First-Person and Bird's-eye View Perspectives

While Experiment I and Experiment II were run independently as within-subject designs, they were conducted in the same environment with the exact procedure applied in both of the studies. Additionally, each prototype mutually shares sonifications as a display mode with the only difference being the viewing point. That is, in Experiment I, participants viewed the environment from a first-person perspective while participants explored the same environment from a bird's-eye view in Experiment II. This makes it possible to compare the performances of the two viewing points with respect to constant and variable errors. There were no significant differences found in constant errors between Experiment I and Experiment II: between the egocentric routes with a difference of 0.1m ± 0.2 SE, t(52) = 0.264, 95% CI [-0.34, 0.49], $p = .793$; between the allocentric routes with a difference of -0.2m ± 0.2 SE, t(52) = -1.092, 95% CI [-0.67, 0.20], $p = .280$. Variable error was significantly lower in Experiment II. That is, between the egocentric routes, there was a significant mean difference of 0.3m ± 0.1 SE, t(26.214) = 3.509, 95% CI [0.12, 0.47], $p = .002$. Between the allocentric routes, there was a significant a mean difference of 0.7m ± 0.3 SE, t(52) = 2.593, 95% CI [0.2, 1.3], $p = .012$.

4.6 Discussion

Along with the hypotheses, in Experiment I and Experiment II, it was expected that the multisensory display modes would increase micro-level navigation accuracy and

precision. In other words, the multisensory modes of Sonification-Kinaesthesia Prototype and Sonification-Tactile Prototype would yield the lowest constant and variable errors (H1). Furthermore, it was predicted that the enhanced micro-level navigation would be the result of multisensory combination, and participants' strategies would mirror how they supplemented their perceptual judgements with multisensory cues (H2).

4.6.1 Performance of Sonification-Kinaesthesia Prototype

In Experiment I, it was found that the unisensory display modes of Sonification-Kinaesthesia Prototype resulted in similar constant errors. This evidenced that a novel sonification display (with 1.55m accuracy) was on a par with an intact kinaesthesia sense (with 1.70m accuracy). Previous research that investigated kinaesthesia or similar sonifications in navigation context also showed similar findings. That is, intact and novel sensory cues could produce equivalent spatial representations for similar navigation accuracy [77,80]. However, even though 62% of participants found the sonification-kinaesthesia mode the easiest, the multisensory mode (with 1.50m accuracy) did not significantly increase overall navigation performance accuracy.

Variable error was significantly different only between the egocentric and allocentric routes completed via kinaesthesia mode with a mean precision difference of 0.80m. This could be explained by the lack of rich sensory information conveyed by the kinaesthesia mode. Participants reported that they counted their steps to calculate distances between targets. While this provided them with a good distance estimate (i.e. egocentric cues), participants reported that understanding the relation between targets in the allocentric route was only 'guess work'. This might consequently explain the significant difference in the variable error. As participants counted their steps to finish navigating at points closer to each other in the egocentric route, hence lower variable error, their guess work randomly dispersed their finishing points in the allocentric route, hence the higher variable error. As the variable error is argued to measure the robustness of perceptual judgement [18], overall, it can be argued that

the unisensory sonification and multisensory sonification-kinaesthesia modes delivered more robust spatial information than the kinaesthesia mode.

Despite the lack of feedback during the experimentation phase, participants showed an overall significant accuracy improvement in unisensory sonification (in egocentric and only marginally in allocentric route, $p = .049$) and multisensory sonification-kinaesthesia (in egocentric route) modes. The effect size of learning in sonification mode, moreover, was relatively larger than that of sonification-kinaesthesia mode. This consequently raises the question of why participants did not show as much improvement and significantly higher performance in the multisensory sonification-kinaesthesia mode as hypothesised. Firstly, it could be because participants did not show improvement in the kinaesthesia mode, which might have pushed back the benefits of multisensory combination. That is, in terms of the switching cost, participants might have been distracted from the contradictory information between unimproved kinaesthetic and improved sonification cues. Alternatively, participants might have needed to have experience with the sonifications for a more extended period of time. This might have reinforced the reliability of the sonification mode for enhanced navigation. This is, however, a less likely explanation since the accuracy of sonification and kinaesthesia were on par. For successful multisensory processing, it is suggested that cues of different origins (e.g. auditory and kinaesthetic) should be reliable [84,95–97]. It might therefore be expected of participants to perform better with the sonification-kinaesthesia mode having acquired more experience with sonifications. This also indicates that learning to take full advantage of the sensory information provided could improve overall cue reliability; however, more research is needed in this area, especially with respect to multisensory combination.

Nonetheless, participants' strategies for using the three display modes clearly revealed that the multisensory mode could be beneficial for micro-level navigation beyond the switching cost. They found the sonification-kinaesthesia mode to be the easiest because it enabled them to enquire both 'distance' and 'angular' information from the environment. This is indeed crucial for navigation as egocentric (e.g. distance cues) and allocentric (e.g. angular cues) information exist in parallel, and combine for

more robust perceptual judgements [14]. Micro-level navigation applications should therefore be able to convey various spatial representations successively via multisensory channels. In Experiment I, however, the sonification-kinaesthesia did not achieve enhanced navigation performance to a significant level. Overall, this might be because the Sonification-Kinaesthesia Prototype was not able to convey multisensory cues, which were robust enough in delivering distinct egocentric and allocentric representations.

4.6.2 Performance of Sonification-Tactile Prototype

In Experiment II, it was found that the unisensory tactile mode (with 1.4m accuracy) of Sonification-Tactile Prototype significantly resulted in lower constant error than the unisensory sonification mode (with 1.7m accuracy) with an average accuracy difference of 0.3m. This evidenced that a tactile cross-modal display could be superior to sonifications. In contrast, it is argued that sonifications with 11,264 auditory pixels (e.g. via The vOICe, the sonification mode of Sonification-Tactile Prototype) would be superior to tactile display modes (e.g. via BrainPort, the tactile mode of Sonification-Tactile Prototype with 400 electro-tactile pixels [108]) because of the higher informational capacity of the sonifications [36]. The findings indicated that the perceived resolution of sensory information might cap the overall informational capacity. That is, even though the sonification mode was higher in informational resolution, participants were able to benefit more from the tactile mode with lower resolution. On the other hand, even though 53% of participants found sonification-tactile mode to be the easiest to use, the multisensory mode (with 1.5m accuracy) did not significantly improve overall navigation performance accuracy. This could be because the information capacity of the sonification and tactile modes were different, thereby decreasing the effectiveness of multisensory combination.

Variable error was significantly different between the egocentric and allocentric routes across all display modes with a mean precision difference of 0.40m. Given that the variable error measures the robustness of perceptual judgements [18], it could be argued that Sonification-Tactile Prototype overall was more suitable for micro-

navigation tasks that did not require turning at an intermediary target. Alternatively, it can be argued that the variable error might not be a good measure of micro-level navigation performed with novel cross-modal displays. Given the effect of learning in both the sonification and tactile modes, participants are expected to end their navigations at points further apart from each other as the trials go on, as a result of improved wayfinding. This would consequently lower the precision of navigation trials, hence increasing the variable error. These findings suggest then that higher variable error is not necessarily a sign of poor performance if there is the effect of learning.

In Experiment II, despite the lack of feedback during the experimentation phase, participants showed an overall significant accuracy improvement in unisensory tactile (in allocentric route) and sonification (in egocentric route) modes. The selective improvements with tactile and sonification modes in allocentric and egocentric routes respectively mirrored how participants utilised their strategies to use each unisensory display mode. That is, sonifications conveyed egocentric information via distance information, and tactile cues created an allocentric representation via angular cues. This is in line with previous research showing that egocentric representations from sonifications can be complemented with allocentric representations from a tactile map [77]. With the sonification-tactile mode, participants further reported to enquire the complementary information and consequently found the multisensory mode to be the easiest. This suggests that sonification and tactile cross-modal displays might have a spatial bias. If this is the case, their multisensory combination might equip the users with robust egocentric and allocentric representations for enhanced navigation.

Despite the reported benefits of the sonification-tactile mode, navigation performance with the multisensory mode did not improve significantly. Participants also found it harder to focus on both cues simultaneously. A plausible explanation for why the multisensory mode did not enhance overall navigation performance could be that there was lack of extensive training and high switching cost. However, the current study found the multisensory performance was not poorly influenced, which indicates that this explanation is unlikely. It might be expected of participants to ignore one of the display modes completely and instead use only the one they relied most on. Only

a few participants, however, reported using this strategy. Instead, most participants strategized to frequently switch between sonifications and tactile cues, suggesting the benefits of multisensory information channels. By doing so, they could have prevented the disadvantages of sensory adaptation without being affected by the switching cost. Nevertheless, participants' reports did not explain the onset of these selective strategies. The strategies participants used were clearly task dependent and future research should investigate the conceptualisation of user strategies while using cross-modal displays.

4.6.3 Performance of Viewing Perspectives

In comparing Experiment I to Experiment II, it was revealed that constant error did not significantly differ between the first-person perspective and bird's-eye view when the sonification mode was used. Replicating this finding between the sonification modes of Sonification-Kinaesthesia and Sonification-Tactile Prototypes further indicated that the viewing perspective did not influence navigation performance. While this might be true for accuracy, participants' strategies highlighted additional variations in how sonifications were used with respect to the viewing perspectives. On contrary to the use of sonifications for allocentric cues (e.g. angular information) in Experiment I, participants explicitly reported that the sonification mode of Sonification-Tactile Prototype in Experiment II conveyed egocentric cues (e.g. distances). This alteration, however, cannot be deduced from the change in perspectives alone. This is because the tactile mode was reported to convey allocentric cues while the sonification mode conveyed egocentric cues. Additionally, in both experiments, participants showed an effect of learning with sonification mode in the egocentric route, suggesting that sonifications might be biased more towards egocentric cues. The variable error from the sonification mode of Sonification-Tactile Prototype, moreover, was significantly lower than that of Sonification-Kinaesthesia Prototype with an average precision difference of 0.5m. Arguably, it may be that the bird's-eye view generally provides a more robust mental map than the first-person perspective.

4.6.4 Future Perspectives

4.6.4.1 Theoretical Implications and Limitations

The discussion of results so far indicates that display modes could be biased towards delivering certain representations of spatial information. The tactile mode was preferred for gaining allocentric cues and kinaesthesia for egocentric cues. This is in line with previous research suggesting that kinaesthetic and tactile sensations lead to egocentric and allocentric representations respectively [77,80]. Sonification mode, on the other hand, showed an alternating pattern. While sonifications were used for collecting egocentric information in Experiment II, they were utilised for gathering allocentric information in Experiment I. This raises the question of whether the spatial representations from sonifications remained in the visual or auditory domain. That is, while visual experience develops allocentric representations [45], it is shown that auditory experience is biased towards egocentric representations [89,103].

The evidence from previous studies examining spatial representations with respect to sonifications is mixed. It is possible to argue that sonifications from cross-modal displays, unlike other environmental auditory sources, carry visual elements [44,72,74]. According to this line of research, it would be argued that sonifications lead to vision-like allocentric representations. On the other hand, spatial representations of sonifications might remain in the auditory domain [43,85]. If so, sonifications would subsequently convey egocentric cues from the environment [77,89,103]. Alternatively, the alteration within the spatial representations that the sonifications created in Experiment I and Experiment II could be task dependent as a consequence of how the environment is learnt [78–80,102,110]. For example, it was evidenced that performance with cross-modal displays was significantly influenced by whether participants were asked to focus on proximal stimulation or distal attributions to explore the environment [90].

In parallel, a recent line of research claims the vertical integration thesis. The thesis argues that cross-modal displays, powered with sensory substitution techniques like the prototypes in the current research, can enable pre-existing capacities of multiple

senses (e.g. both the substituting and substituted) for the given task and alter users' strategies accordingly [4,7,8,24,25]. This suggests that sonifications might have evoked both vision-like and audition-like representations, hence allocentric and egocentric respectively, in Experiment I and Experiment II. It is therefore important to investigate what the primary deciding factors are for users to allocate their strategies. Participants, for example, reported that they actively moved the camera in Experiment I and relocated the intra-oral display in Experiment II. This enabled them active movement capabilities with the sonification mode in Experiment I and tactile mode in Experiment II. Incidentally, participants also acquired allocentric representations in these conditions where they were able to actively explore the environment. Future research should therefore consider the role of the stationary and moving observers, and passive and active movement in designing cross-modal displays with multisensory modes [90]. Identifying their role with respect to user strategies would further reveal how the benefits of multisensory combination can be efficiently maximised.

Multisensory combination leads to robust perceptual judgements when the cues complement each other [71]. Multisensory integration, on the other hand, requires redundant cues [87]. This suggestively points out that the two multisensory processes might be dependent on a complementary-redundant cue spectrum. Nevertheless, what happens when some cues are redundant, and others are complementary is unknown. It could be the case that the two processes would compete for cognitive resources when redundant and complementary cues are provided simultaneously. This would deprive the overall multisensory response quality. In light of the vertical integration thesis, this consideration is also relevant to designing cross-modal displays with multisensory modes. That is, the cross-modal displays can carry both redundant (i.e. from the substituted sensory source) and complementary (i.e. from the substituting sensory source) cues.

In participants' strategies, it was clear that they were searching for complementary egocentric and allocentric information for creating a robust spatial representation regardless of the display mode. They found it to be the easiest with the multisensory

modes. The fact that participants also performed well with unisensory modes (e.g. with the tactile mode), however, indicates that unisensory modes were able to convey both egocentric and allocentric cues as well. This is suggestive of the redundancy of cues when they were presented in the multisensory mode. Moreover, the redundant and complimentary nature of cross-modal cues can be at varying degrees as participants reported to find the sonifications and tactile cues to deliver higher horizontal and vertical spatial resolutions respectively. Future research should therefore target how the axis speciality of sonification and tactile modes could be combined in terms of providing robust egocentric and allocentric information. In this way, it is suggested that multisensory combination would improve overall navigation performance. This also requires further investigation of redundancy with respect to perception with cross-modal displays, and how the relation between complementary and redundant cues might influence multisensory processing.

4.6.4.2 Practical Implications and Limitations

Sonification-kinaesthesia mode of Sonification-Kinaesthesia Prototype and tactile mode of Sonification-Tactile Prototype resulted in the lowest average navigation accuracy of 1.5m and 1.4m respectively. In other words, they were able to bring the last 10-metres problem down to approximately 1.5m. To our best knowledge, how frequently users use web mapping services on their mobile phones during navigation to confirm their location and how long it takes for them to inspect their remaining route are unknown. In the current research, participants were able to learn an environment within 10 seconds and reach their final target within the proximity of 1.5m. This presents a successful use case of cross-modal displays for micro-level navigation. Participants should in theory perform better if they also received real-time sensory information during navigation. To incorporate this in the experimental procedure, time to navigate a route could also be studied along with accuracy and precision errors. These encourage new areas of research as well as the development of cross-modal displays with multisensory modes for inclusive assisted navigation applications.

For developing multisensory display modes, one direction to consider is to investigate the necessary threshold of temporal congruency of cross-modal cues for multisensory combination. Participants who used the multisensory mode of Sonification-Tactile Prototype, for example, reported that they strategized to focus on auditory and tactile cues by switching their attention from one to the other repeatedly. The current results do not give an insight into whether such a strategy was the cause why the multisensory sonification-tactile mode did not outperform the tactile mode. However, the fact that participants did not ignore one of the cues still suggests that multisensory information channels were looked for richer spatial representations. If this is true, cross-modal displays with a multisensory mode that deliver cues in an alternating sequence might perform better than the ones that deliver multisensory cues simultaneously. Indeed, recent research showed that even periodical switching between cues of different origins could improve navigation performance by counteracting the disadvantages of sensory adaptation and switching cost [52]. Temporally misaligning cues, after a threshold, is known to deprive the quality of the multisensory response, especially in the case of multisensory integration [97,98]. The threshold of such incongruencies in multisensory feedback and their possible benefits, however, are not thoroughly tested in the use cases of cross-modal displays with multisensory modes, which empower multisensory combination.

In relation to this, an additional direction to explore is through examining ways to drive forward users' selective and adaptive strategies with display modes. For example, even though spatial updating is evidenced to be in co-existence between egocentric and allocentric representations [14], the frequency of the updating is unknown. In other words, users may require different types of spatial representations in a temporal sequence rather than their constant presentation. In this way, navigation applications could provide allocentric representations instead of egocentric views at critical points in micro-level navigation, or vice versa. For example, allocentric information can be provided to the user when they are about to make a turn. Similarly, the viewing point from bird's-eye to first-person can be changed depending on the users' needs. Along these lines, the current research further evidenced that cross-modal displays could be better at providing axis special resolution (e.g. sonifications

for higher horizontal resolution and tactile cues for higher vertical resolution). The axis speciality of cross-modal displays can be coupled with alternating cue presentation to deliver enhanced navigation capabilities by studying when the users need different kinds of spatial representations. This is in line with previous research, which suggested that horizontal and vertical navigation form a more robust spatial representation of the environment [102]. By addressing these limitations and future research directions, task adaptive cross-modal cue selection can be further implemented via multisensory combination techniques at users' manual control or at algorithmic level for inclusive navigation applications [51,53].

Alternatively, micro-level navigation applications could deliver cross-modal sonifications and tactile cues when users are needed to be rerouted or when they are lost. Feedback can therefore be task adaptive and conveyed accordingly, minimising the interaction time. This would be similar to using thermal feedback as means of communicating spatial information [10,67,101,107]. These attempts to cross-modally pair spatial information with thermal feedback resemble to the children's game where the location of a hidden object is cued with either 'Hot' or 'Cold'. However, such unisensory feedback lacks the dimensional complexity to represent both egocentric and allocentric information for successful locomotion and wayfinding. Instead, by prompting adaptive positional sonifications and/or tactile cues in real-time when the route is 'cold' can lead to successful micro-navigation. Reducing the frequency of cues can further eliminate the need of training with cross-modal displays for seamless, intuitive and inclusive user experience.

4.7 Conclusion

Being freely mobile gives us independence, and enhances our quality of life and social connectivity. This fundamentally places a vital role on inclusive navigation applications and encourages their widespread adoption by various user profiles. However, dependency on unisensory display modes for interacting with the digital world inherently becomes an obstacle as they unavoidably exclude certain user profiles and

use cases. Examining technologies that provide alternative display modes for different users is paramount for building an inclusive society. The current research shows how non-visual cross-modal displays with unisensory and multisensory modes can be prototyped and examined with a mixed study approach in enhancing assisted navigation applications, especially in the context of micro-navigation. The advantages and disadvantages of sonifications, tactile cues, and their multisensory combination are demonstrated in targeting the last 10-metres of navigation. Overall, the tactile cross-modal display prototype is shown to bring the last 10-metres problem down to 1.5 metres. The current research evidences how cross-modal displays utilising sensory substitution techniques lead to task-dependent spatial representations, which consequently alters users' strategies in navigation. The findings further support how egocentric and allocentric representations co-exist, and are sought by users for successful locomotion and wayfinding. It is discussed that assistive navigation applications should be examined more holistically by considering the cognitive processes that connect micro- and macro- level navigation abilities of users. In future, how egocentric and allocentric representations could be delivered to the users with multisensory combination in an alternating temporal sequence or simultaneous delivery should be addressed for developing inclusive navigation applications.

4.8 References

1. Mamta Agiwal, Abhishek Roy, and Navrati Saxena. 2016. Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials* 18, 3: 1617–1655. https://doi.org/10.1109/COMST.2016.2532458
2. Jeffrey G. Andrews, Stefano Buzzi, Wan Choi, Stephen V. Hanly, Angel Lozano, Anthony C. K. Soong, and Jianzhong Charlie Zhang. 2014. What Will 5G Be? *IEEE Journal on Selected Areas in Communications* 32, 6: 1065–1082. https://doi.org/10.1109/JSAC.2014.2328098
3. Flora M. Antunes and Manuel S. Malmierca. 2014. An Overview of Stimulus-Specific Adaptation in the Auditory Thalamus. *Brain Topography* 27, 4: 480–499. https://doi.org/10.1007/s10548-013-0342-6
4. Gabriel Arnold, Jacques Pesnot-Lerousseau, and Malika Auvray. 2017. Individual Differences in Sensory Substitution. *Multisensory Research* 30, 6: 579–600. https://doi.org/10.1163/22134808-00002561
5. ATLAS.ti. 2019. ATLAS.ti: The Qualitative Data Analysis & Research Software. *ATLAS.ti*. Retrieved from https://atlasti.com/
6. AutoDesk. 2019. AutoCAD. *AutoDesk*. Retrieved from https://www.autodesk.com/products/autocad/overview
7. Malika Auvray and Mirko Farina. 2017. Patrolling the Boundaries of Synaesthesia. In *Synaesthesia: Philosophical & Psychological Challenges*, O Deroy (ed.). Oxford University Press, Oxford, 248–274.
8. Malika Auvray and Erik Myin. 2009. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science* 33, 6: 1036–1058. https://doi.org/10.1111/j.1551-6709.2009.01040.x
9. Paul Bach-y-Rita and Stephen W. Kercel. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences* 7, 12: 541–546. https://doi.org/10.1016/J.TICS.2003.10.013
10. Jan Balata, Katerina Prazakova, Anna Kutíková, and Miroslav Macik. 2013. Quido: Arcade Game with Thermo-Haptic Feedback. In *Cognitive Infocommunications*.
11. David H. Brainard. 1997. The Psychophysics Toolbox. *Spatial Vision* 10, 4: 433–436. https://doi.org/10.1163/156856897X00357
12. David Brown, Tom Macpherson, and Jamie Ward. 2011. Seeing with Sound? Exploring Different Characteristics of a Visual-to-Auditory Sensory Substitution Device. *Perception* 40, 9: 1120–1135. https://doi.org/10.1068/p6952
13. Heinrich H. Bülthoff and Hanspeter A. Mallot. 1988. Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A* 5, 10: 1749. https://doi.org/10.1364/JOSAA.5.001749
14. Neil Burgess. 2006. Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences* 10, 12: 551–557. https://doi.org/10.1016/J.TICS.2006.10.005
15. Austin McRae Butts. 2015. Enhancing the Perception of Speech Indexical Properties of Cochlear Implants through Sensory Substitution. Arizona State University.
16. Gemma A. Calvert, Charles Spence, and Barry E. Stein. 2004. The Handbook of Multisensory Processing.
17. Marco Centenaro, Lorenzo Vangelista, Andrea Zanella, and Michele Zorzi. 2016. Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios. *IEEE Wireless Communications* 23, 5: 60–67. https://doi.org/10.1109/MWC.2016.7721743
18. Ken Cheng, Sara J. Shettleworth, Janellen Huttenlocher, and John J. Rieser. 2007. Bayesian integration of spatial information. *Psychological Bulletin* 133, 4: 625–637. https://doi.org/10.1037/0033-2909.133.4.625
19. Woon Chin, Zhong Fan, and Russell Haines. 2014. Emerging technologies and research challenges for 5G wireless networks. *IEEE Wireless Communications* 21, 2: 106–112. https://doi.org/10.1109/MWC.2014.6812298
20. Xuerong Cui, Thomas Aaron Gulliver, Juan Li, and Hao Zhang. 2016. Vehicle Positioning Using 5G Millimeter-Wave Systems. *IEEE Access* 4: 6964–6973. https://doi.org/10.1109/ACCESS.2016.2615425

21. Ruth C. Dalton, Christoph Hölscher, and Daniel R. Montello. 2019. Wayfinding as a Social Activity. *Frontiers in Psychology* 10: 142. https://doi.org/10.3389/fpsyg.2019.00142
22. Armin Dammann, Ronald Raulefs, and Siwei Zhang. 2015. On prospects of positioning in 5G. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, 1207–1213. https://doi.org/10.1109/ICCW.2015.7247342
23. David Eagleman. 2015. Can we create new senses for humans? *TED*. Retrieved from https://www.ted.com/talks/david_eagleman_can_we_create_new_senses_for_humans
24. Ophelia Deroy and Malika Auvray. 2012. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00457
25. Ophelia Deroy and Malika Auvray. 2014. A Crossmodal Perspective on Sensory Substitution. In *Perception and Its Modalities*. Oxford University Press, 327–349. https://doi.org/10.1093/acprof:oso/9780199832798.003.0014
26. D. M. Eagleman, S. D. Novich, D. Goodman, A. Sahoo, and M. Perotta. 2017. Method and system for providing adjunct sensory information to a user. Retrieved from https://patents.google.com/patent/US10198076B2/en
27. Marc O. Ernst and Heinrich H. Bülthoff. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences* 8, 4: 162–169. https://doi.org/10.1016/J.TICS.2004.02.002
28. A S Etienne, R Maurer, and V Séguinot. 1996. Path integration in mammals and its interaction with visual landmarks. *The Journal of experimental biology* 199, Pt 1: 201–9.
29. Ariane S. Etienne and Kathryn J. Jeffery. 2004. Path integration in mammals. *Hippocampus* 14, 2: 180–192. https://doi.org/10.1002/hipo.10173
30. Nicholas A. Giudice and Jerome D. Tietz. 2008. Learning with Virtual Verbal Displays: Effects of Interface Fidelity on Cognitive Map Development. In *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*. Springer Berlin Heidelberg, Berlin, Heidelberg, 121–137. https://doi.org/10.1007/978-3-540-87601-4_11
31. J Goodman, SA Brewster, and P Gray. 2005. How can we best use landmarks to support older people in navigation? *Behaviour & Information Technology* 24, 1: 3–20. https://doi.org/10.1080/01449290512331319021
32. Charles Gouin-Vallerand and Simon Rousseau. 2019. An indoor navigation platform for seeking Internet of Things devices in large indoor environment. In *Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good - GoodTechs '19*, 108–113. https://doi.org/10.1145/3342428.3342652
33. Kalanit Grill-Spector, Richard Henson, and Alex Martin. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10, 1: 14–23. https://doi.org/10.1016/J.TICS.2005.11.006
34. A. Gupta and R. K. Jha. 2015. A Survey of 5G Network: Architecture and Emerging Technologies. *IEEE Access* 3: 1206–1232. https://doi.org/10.1109/ACCESS.2015.2461602
35. Yoram Gutfreund. 2012. Stimulus-specific adaptation, habituation and change detection in the gaze control system. *Biological Cybernetics* 106, 11–12: 657–668. https://doi.org/10.1007/s00422-012-0497-3
36. Alastair Haigh, David J. Brown, Peter Meijer, and Michael J. Proulx. 2013. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00330
37. Badis Hammi, Rida Khatoun, Sherali Zeadally, Achraf Fayad, and Lyes Khoukhi. 2018. IoT technologies for smart cities. *IET Networks* 7, 1: 1–13. https://doi.org/10.1049/iet-net.2017.0163
38. C. Harrison, B. Eckman, R. Hamilton, P. Hartswick, J. Kalagnanam, J. Paraszczak, and P. Williams. 2010. Foundations for Smarter Cities. *IBM Journal of Research and Development* 54, 4: 1–16. https://doi.org/10.1147/JRD.2010.2048257
39. Cristy. Ho and Charles. Spence. 2008. *The Multisensory Driver : Implications for Ergonomic Car Interface Design.* CRC Press.
40. E.E. Hoggan and S.A. Brewster. 2006. Crossmodal Interaction with Mobile Devices. In *Visual Languages and Human-Centric Computing (VL/HCC'06)*, 234–235. https://doi.org/10.1109/VLHCC.2006.18
41. Eve Hoggan and Stephen Brewster. 2007. Designing audio and tactile crossmodal icons for

mobile devices. In *Proceedings of the ninth international conference on Multimodal interfaces - ICMI '07*, 162. https://doi.org/10.1145/1322192.1322222
42. Eve Hoggan, Topi Kaaresoja, Pauli Laitinen, and Stephen Brewster. 2008. Crossmodal congruence. In *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*, 157. https://doi.org/10.1145/1452392.1452423
43. Nicholas. Humphrey. 1992. *A history of the mind*. Chatto & Windus, London.
44. Susan Hurley and Alva Noë. 2003. Neural Plasticity and Consciousness. *Biology & Philosophy* 18, 1: 131–168. https://doi.org/10.1023/A:1023308401356
45. Tina Iachini, Gennaro Ruggiero, and Francesco Ruotolo. 2014. Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behavioural Brain Research* 273: 73–81. https://doi.org/10.1016/J.BBR.2014.07.032
46. IBM. 2019. SPSS Software. *IBM*. Retrieved from https://www.ibm.com/analytics/spss-statistics-software
47. Lynette A Jones, Jacquelyn Kunkel, and Erin Piateski. 2009. Vibrotactile Pattern Recognition on the Arm and Back. *Perception* 38, 1: 52–68. https://doi.org/10.1068/p5914
48. K.A. Kaczmarek, J.G. Webster, P. Bach-y-Rita, and W.J. Tompkins. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38, 1: 1–16. https://doi.org/10.1109/10.68204
49. Amy A. Kalia, Paul R. Schrater, and Gordon E. Legge. 2013. Combining Path Integration and Remembered Landmarks When Navigating without Vision. *PLoS ONE* 8, 9: e72170. https://doi.org/10.1371/journal.pone.0072170
50. R M Klein. 1977. Attention and visual dominance: a chronometric analysis. *Journal of experimental psychology. Human perception and performance* 3, 3: 365–78. https://doi.org/10.1037//0096-1523.3.3.365
51. Kyle Kotowick and Julie Shah. 2017. Intelligent Sensory Modality Selection for Electronic Supportive Devices. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, 55–66. https://doi.org/10.1145/3025171.3025228
52. Kyle Kotowick and Julie Shah. 2018. Modality Switching for Mitigation of Sensory Adaptation and Habituation in Personal Navigation Systems. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*, 115–127. https://doi.org/10.1145/3172944.3172980
53. Kyle Kotowick and Julie Shah. 2018. Effects of an Adaptive Modality Selection Algorithm for Navigation Systems. In *The 31st Annual ACM Symposium on User Interface Software and Technology - UIST '18*, 543–556. https://doi.org/10.1145/3242587.3242610
54. Simon Lacey and K. Sathian. 2014. Visuo-haptic multisensory object recognition, categorization, and representation. *Frontiers in Psychology* 5. https://doi.org/10.3389/fpsyg.2014.00730
55. C. Lenay, S. Canu, and P. Villon. 1997. Technology and perception: the contribution of sensory substitution systems. In *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*, 44–53. https://doi.org/10.1109/CT.1997.617681
56. Charles Lenay, Olivier Gapenne, Sylvain Hanneton, Catherine Marque, and Christelle Genouëlle. 2003. SENSORY SUBSTITUTION: LIMITS AND PERSPECTIVES. In *Touching for Knowing: Cognitive Psychology of Haptic Manual Perception*, Y Hatwell, A Streri and E Gentaz (eds.). John Benjamins Publishing Company , Amsterdam, Netherlands, 275–292
57. Ying Liu, Xiufang Shi, Shibo He, and Zhiguo Shi. 2017. Prospective Positioning Architecture and Technologies in 5G Networks. *IEEE Network* 31, 6: 115–121. https://doi.org/10.1109/MNET.2017.1700066
58. Diana Löffler, Robert Tscharn, and Jörn Hurtienne. 2018. Multimodal Effects of Color and Haptics on Intuitive Interaction with Tangible User Interfaces. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '18*, 647–655. https://doi.org/10.1145/3173225.3173257
59. Shachar Maidenbaum, Roni Arbel, Galit Buchs, Shani Shapira, and Amir Amedi. 2014. Vision through other senses: Practical use of Sensory Substitution devices as assistive technology for visual rehabilitation. In *22nd Mediterranean Conference on Control and Automation*, 182–187. https://doi.org/10.1109/MED.2014.6961368
60. Mathworks. 2019. MATLAB. *Mathworks*. Retrieved from https://www.mathworks.com/products/matlab.html
61. P.B.L. Meijer. 1992. An experimental system for auditory image representations. *IEEE*

Transactions on Biomedical Engineering 39, 2: 112–121. https://doi.org/10.1109/10.121642
62. Krista Merry and Pete Bettinger. 2019. Smartphone GPS accuracy study in an urban environment. *PLOS ONE* 14, 7: e0219890. https://doi.org/10.1371/journal.pone.0219890
63. Esmond Mok, Guenther Retscher, and Chen Wen. 2012. Initial test on the use of GPS and sensor data of modern smartphones for vehicle tracking in dense high rise environments. In *2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS)*, 1–7. https://doi.org/10.1109/UPINLBS.2012.6409789
64. Daniel R. Montello. 2005. Navigation. In *The Cambridge Handbook of Visuospatial Thinking*, Priti Shah and Akira Miyake (eds.). Cambridge University Press, Cambridge, 257–294. https://doi.org/10.1017/CBO9780511610448.008
65. Daniel Montello and Corina Sas. 2006. Human Factors of Wayfinding in Navigation. In *International Encyclopedia of Ergonomics and Human Factors, Second Edition - 3 Volume Set* (2nd ed.), Waldemar Karwowski (ed.). CRC Press. https://doi.org/10.1201/9780849375477.ch394
66. Weimin Mou, Timothy P. McNamara, Christine M. Valiquette, and Björn Rump. 2004. Allocentric and Egocentric Updating of Spatial Memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30, 1: 142–157. https://doi.org/10.1037/0278-7393.30.1.142
67. Mutsuhiro Nakashige, Minoru Kobayashi, Yuriko Suzuki, Hidekazu Tamaki, and Suguru Higashino. 2009. "Hiya-Atsu" media: augmenting digital media with temperature. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, 3181. https://doi.org/10.1145/1520340.1520453
68. Marko Nardini, Peter Jones, Rachael Bedford, and Oliver Braddick. 2008. Development of Cue Integration in Human Navigation. *Current Biology* 18, 9: 689–693. https://doi.org/10.1016/J.CUB.2008.04.021
69. National Research Council. 2008. *Emerging Cognitive Neuroscience and Related Technologies*. National Academies Press, Washington, D.C. https://doi.org/10.17226/12177
70. Paolo Neirotti, Alberto De Marco, Anna Corinna Cagliano, Giulio Mangano, and Francesco Scorrano. 2014. Current trends in Smart City initiatives: Some stylised facts. *Cities* 38: 25–36. https://doi.org/10.1016/j.cities.2013.12.010
71. Fiona N. Newell, Marc O. Ernst, Bosco S. Tjan, and Heinrich H. Bülthoff. 2001. Viewpoint Dependence in Visual and Haptic Object Recognition. *Psychological Science* 12, 1: 37–42. https://doi.org/10.1111/1467-9280.00307
72. Alva. Noë. 2004. *Action in perception*. MIT Press.
73. Scott D. Novich and David M. Eagleman. 2014. A vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In *2014 IEEE Haptics Symposium (HAPTICS)*, 1–1. https://doi.org/10.1109/HAPTICS.2014.6775558
74. J. K. O'Regan. 2011. *Why red doesn't sound like a bell : understanding the feel of consciousness*. Oxford University Press.
75. Marianna Obrist, Elia Gatti, Emanuela Maggioni, Chi Thanh Vi, and Carlos Velasco. 2017. Multisensory Experiences in HCI. *IEEE MultiMedia* 24, 2: 9–13. https://doi.org/10.1109/MMUL.2017.33
76. Sharon Oviatt and Sharon. 1999. Ten myths of multimodal interaction. *Communications of the ACM* 42, 11: 74–81. https://doi.org/10.1145/319382.319398
77. Achille Pasqualotto and Tayfun Esenkaya. 2016. Sensory Substitution: The Spatial Updating of Auditory Scenes "Mimics" the Spatial Updating of Visual Scenes. *Frontiers in Behavioral Neuroscience* 10. https://doi.org/10.3389/fnbeh.2016.00079
78. Achille Pasqualotto, Ciara M. Finucane, and Fiona N. Newell. 2013. Ambient visual information confers a context-specific, long-term benefit on memory for haptic scenes. *Cognition* 128, 3: 363–379. https://doi.org/10.1016/J.COGNITION.2013.04.011
79. Achille Pasqualotto and Michael J. Proulx. 2013. The study of blindness and technology can reveal the mechanisms of three-dimensional navigation. *Behavioral and Brain Sciences* 36, 5: 559–560. https://doi.org/10.1017/S0140525X13000496
80. Achille Pasqualotto, Mary Jane Spiller, Ashok S. Jansari, and Michael J. Proulx. 2013. Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research* 236: 175–179. https://doi.org/10.1016/J.BBR.2012.08.042
81. Aftab E. Patla, Stephen D. Prentice, and Lilian T. Gobbi. 1996. Visual Control of Obstacle

Avoidance During Locomotion: Strategies in Young Children, Young and Older Adults. *Advances in Psychology* 114: 257–277. https://doi.org/10.1016/S0166-4115(96)80012-4

82. Aftab E. Patla and Joan N. Vickers. 1997. Where and when do we look as we approach and step over an obstacle in the travel path? *NeuroReport* 8, 17: 3661–3665. https://doi.org/10.1097/00001756-199712010-00002
83. Diane Pecher, René Zeelenberg, and Lawrence W. Barsalou. 2003. Verifying Different-Modality Properties for Concepts Produces Switching Costs. *Psychological Science* 14, 2: 119–124. https://doi.org/10.1111/1467-9280.t01-1-01429
84. Thomas J. Perrault, J. William Vaughan, Barry E. Stein, and Mark T. Wallace. 2003. Neuron-Specific Response Characteristics Predict the Magnitude of Multisensory Integration. *Journal of Neurophysiology* 90, 6: 4022–4026. https://doi.org/10.1152/jn.00494.2003
85. Jesse J. Prinz. 2006. Putting the brakes on enactive perception. *PSYCHE: An Interdisciplinary Journal of Research On Consciousness*, 12
86. Michael J. Proulx and A. Harder. 2008. Sensory substitution: Visual-to-auditory sensory substitution devices for the blind. *Tijdschrift voor Ergonomie* 6, 33.
87. Marieke Rohde, Loes C J van Dam, and Marc Ernst. 2016. Statistically Optimal Multisensory Cue Integration: A Practical Tutorial. *Multisensory research* 29, 4–5: 279–317.
88. Peter J. Rousseeuw and Katrien Van Driessen. 1999. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* 41, 3: 212–223. https://doi.org/10.1080/00401706.1999.10485670
89. Victor R. Schinazi, Tyler Thrash, and Daniel-Robert Chebat. 2016. Spatial navigation by congenitally blind individuals. *Wiley Interdisciplinary Reviews: Cognitive Science* 7, 1: 37–58. https://doi.org/10.1002/wcs.1375
90. Joshua H Siegle and William H Warren. 2010. Distal Attribution and Distance Perception in Sensory Substitution. *Perception* 39, 2: 208–223. https://doi.org/10.1068/p6366
91. Daniel J. Simons and Ranxiao Frances Wang. 1998. Perceiving Real-World Viewpoint Changes. *Psychological Science* 9, 4: 315–320. https://doi.org/10.1111/1467-9280.00062
92. K E Skouby and P Lynggaard. 2014. Smart home and smart city solutions enabled by 5G, IoT, AAI and CoT services. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 874–878. https://doi.org/10.1109/IC3I.2014.7019822
93. Charles Spence, Michael E. R. Nicholls, and Jon Driver. 2001. The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics* 63, 2: 330–336. https://doi.org/10.3758/BF03194473
94. Sharmila Sreetharan and Michael Schutz. 2019. Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope. *Multimodal Technologies and Interaction* 3, 1: 4. https://doi.org/10.3390/mti3010004
95. T. R. Stanford. 2005. Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience* 25, 28: 6499–6508. https://doi.org/10.1523/JNEUROSCI.5095-04.2005
96. Terrence R. Stanford and Barry E. Stein. 2007. Superadditivity in multisensory integration: putting the computation in context. *NeuroReport* 18, 8: 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315
97. B E Stein and M T Wallace. 1996. Comparisons of cross-modality integration in midbrain and cortex. *Progress in brain research* 112: 289–99. https://doi.org/10.1016/s0079-6123(08)63336-1
98. Barry E. Stein. 1998. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research* 123, 1–2: 124–135. https://doi.org/10.1007/s002210050553
99. Barry E. Stein, David Burr, Christos Constantinidis, Paul J. Laurienti, M. Alex Meredith, Thomas J. Perrault, Ramnarayan Ramachandran, Brigitte Röder, Benjamin A. Rowland, K. Sathian, Charles E. Schroeder, Ladan Shams, Terrence R. Stanford, Mark T. Wallace, Liping Yu, and David J. Lewkowicz. 2010. Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience* 31, 10: 1713–1720. https://doi.org/10.1111/j.1460-9568.2010.07206.x
100. L. Tcheang, H. H. Bulthoff, and N. Burgess. 2011. Visual influence on path integration in darkness indicates a multimodal representation of large-scale space. *Proceedings of the National Academy of Sciences* 108, 3: 1152–1157. https://doi.org/10.1073/pnas.1011843108

101. Jordan Tewell, Jon Bird, and George R. Buchanan. 2017. The Heat is On: A Temperature Display for Conveying Affective Feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 1756–1767. https://doi.org/10.1145/3025453.3025844
102. Guillaume Thibault, Achille Pasqualotto, Manuel Vidal, Jacques Droulez, and Alain Berthoz. 2013. How does horizontal and vertical navigation influence spatial memory of multifloored environments? *Attention, Perception, & Psychophysics* 75, 1: 10–15. https://doi.org/10.3758/s13414-012-0405-x
103. Stephen M. Town, W. Owen Brimijoin, and Jennifer K. Bizley. 2017. Egocentric and allocentric representations in auditory cortex. *PLOS Biology* 15, 6: e2001878. https://doi.org/10.1371/journal.pbio.2001878
104. Narseo Vallina-Rodriguez, Jon Crowcroft, Alessandro Finamore, Yan Grunenberger, and Konstantina Papagiannaki. 2013. When assistance becomes dependence: characterizing the costs and inefficiencies of A-GPS. *ACM SIGMOBILE Mobile Computing and Communications Review* 17, 4: 3–14. https://doi.org/10.1145/2557968.2557970
105. Sabine Verboven and Mia Hubert. 2005. LIBRA: a MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems* 75, 2: 127–136. https://doi.org/10.1016/J.CHEMOLAB.2004.06.003
106. Stephan von Watzdorf and Florian Michahelles. 2010. Accuracy of positioning data on smartphones. In *Proceedings of the 3rd International Workshop on Location and the Web - LocWeb '10*, 1–4. https://doi.org/10.1145/1899662.1899664
107. Reto Wettach, Christian Behrens, Adam Danielsson, and Thomas Ness. 2007. A thermal information display for mobile applications. In *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services - MobileHCI '07*, 182–185. https://doi.org/10.1145/1377999.1378004
108. Wicab. Wicab, Inc. | United States | BrainPort Technologies. Retrieved from https://www.wicab.com/wicab-inc
109. Jan M. Wiener, Alain Berthoz, and Thomas Wolbers. 2011. Dissociable cognitive mechanisms underlying human path integration. *Experimental Brain Research* 208, 1: 61–71. https://doi.org/10.1007/s00221-010-2460-7
110. T. Wolbers and Christian Büchel. 2005. Dissociable Retrosplenial and Hippocampal Contributions to Successful Formation of Survey Representations. *Journal of Neuroscience* 25, 13: 3333–3340. https://doi.org/10.1523/JNEUROSCI.4705-04.2005
111. Ping Zhang, Jian Lu, Yan Wang, and Qiao Wang. 2017. Cooperative localization in 5G networks: A survey. *ICT Express* 3, 1: 27–32. https://doi.org/10.1016/J.ICTE.2017.03.005

Reflections V

"From a networked perspective, these relationships among people, ideas and objects are the technology"

Andrew Hargadon[3]

The previous study exemplified how sonification and tactile cross-modal display modes could be prototyped and evaluated with a mixed methods design in developing inclusive navigation applications. The first experiment established a baseline for navigation performance with a novel form of interaction (sonifications) in comparison to that of intact kinaesthesia. The results found that the performance of sonification mode was on a par with the kinaesthesia mode. It was further suggested that the perception of space with sonification and kinaesthesia modes might be biased towards its allocentric and egocentric representations respectively. The second experiment found that the tactile mode resulted in the highest navigation performance. Unlike the allocentric use of sonification mode in this first experiment, it was used for gaining an egocentric representation of the space in the second experiment. The tactile mode, however, conveyed allocentric cues. The allocentric use of sonification mode in Experiment I and tactile mode in Experiment II was therefore preliminarily attributed to the ability to actively explore the environment. The axis-speciality, which was earlier discussed in Chapter II and Chapter III, was also evident in the previous chapter. The majority of users reported that tactile cross-modal mode conveyed higher perceived resolution on the vertical axis and the auditory mode provided higher perceived resolution on the horizontal axis. These further evidenced that perception with cross-modal displays utilise multiple senses, thereby altering users' strategies. Overall, the previous study contributed to the testing of the vertical integration thesis in a multisensory context.

[3] From Technology Brokering and The Pursuit of Innovation, 2002, p8.

The previous experiments also suggested that users were not negatively affected by sensory adaptation and switching cost when they used the multisensory modes. Contrary to previous literature, this indicates that multisensory displays can indeed be beneficial to enrich human-computer interactions. Nevertheless, the multisensory modes did not improve overall performance. Despite this, users' strategies revealed that users looked for cues that carried both egocentric and allocentric representations of the environment for successful navigation. For this reason, the multisensory modes were reported to be the easiest. This raises the question of why the multisensory modes did not improve navigation performance. Chapter I argued against this with how design can enable or disable our physical and cognitive capabilities. In this respect, the previous study suggested that the difference in informational capacity of sonification and tactile modes, for example, might be the reason why their multisensory combination did not improve overall navigation performance. In Chapter III, where the multisensory mode of Cross-Modal Box significantly resulted in the highest object recognition performance, the informational capacity of the auditory and tactile modes was equalised at 400 pixels. The current state of knowledge into multisensory combination cannot yet explain how the informational capacity of multisensory cues, and hence their reliabilities with respect to learning, might influence multisensory response quality. More research is therefore needed in this area. An alternative explanation could be that navigation is a more complex task than object recognition. If this is the case, additional considerations will be necessary when developing future inclusive displays.

Overall, Chapter IV explored the potential of inclusive applications of cross-modal display modes to tackle the last 10 metres of navigation. It also showed that cross-modal display modes could guide users to a desired destination with an approximate accuracy of 1.5 metres. Here it is important to note that the last 10-metres problem is relevant to the technological infrastructure and how well navigation applications could interact with users. It was therefore asserted that sensory substitution techniques could display spatial cues from future micro-level navigation applications. The previous chapter also argued that micro- and macro- level navigation, hence locomotion and wayfinding respectively, cannot be completely separated from each

other. That is, inclusive navigation applications should provide users with complementary egocentric and allocentric feedback. This could be achieved via cross-modal displays whose multisensory modes convey complementary spatial representations with multisensory combination. Accordingly, the previous chapter also identified future research directions and possible ways of presenting (e.g. alternating or simultaneous delivery of multisensory modes) spatial cues. In this way, users could benefit from inclusive navigation applications by customising the sensory channels they interact with.

The last 10-metre problem in the context of micro-level navigation was initially proposed by Stephen Brewster as a design challenge in Accessible Infrastructures for the Mobility and Education of Blind People Workshop. Here we have the opportunity to deeply thank Stephen for his inspirational and insightful discussion of the topic. We also thank the Newton Fund for sponsoring this workshop, and Marion Hersh from Glasgow University, Scotland, Alejandro Garcia from Universidade do Vale de Itajai (Univali), Brazil, and everyone involved in organising the workshop.

Closing Summary

> "If the body is sufficiently tuned to the environment, you don't need cognition…the circular causality of us casually embedded in a world through an embodiment of our brain and its cognition"
>
> Karl Friston

In Chapter I, the inclusive design mindset was reviewed with a multidisciplinary perspective. It was argued that the motivation for inclusion would promote the applications of sensory substitution techniques to mainstream technologies. This would consequently improve their implementation as assistive devices and also expand their mainstream adoption. This is achievable by following two pathways. The first pathway argued that the inclusion mindset would unify the research and development resources towards the applications of sensory substitution techniques. As a result, the understanding and knowledge of sensory substitution would expand and its practical applications would be evaluated in various use cases, including rehabilitation purposes. The second pathway argued that the development of cross-modal display modes would democratise our screen dependency equally with other senses. This is essentially different than the implementation of sensory substitution techniques to compensate for the missing sensory forms. As sensory substitution phenomena grant access to information independent of the sensory origin, cross-modal displays with unisensory and multisensory modes would give us the opportunity to customise how we interact with the same technology. In this way, devices can be made usable and accessible by a wider range of people by default. Overall, the present thesis supports the idea that sensory substitution techniques should be applied with a supplementary framework, as opposed to an assistive framework [12–14]. Accordingly, it contributes an inclusive design mindset to the potential applications of sensory substitution techniques as cross-modal displays for all of us.

Chapter II, Chapter III and Chapter IV explored different applications of sensory substitution techniques, in the context of HCI, with an inclusive mindset. One of the mutual theoretical research themes the chapters supported was the vertical integration hypothesis [1–3,6,7], which suggests that perception with sensory substitution utilises the pre-existing capabilities of multiple senses. The current thesis extends this research further by contributing to its investigation in a multisensory context. It moreover evaluates the implications of this in inclusive display development and how cross-modal displays with multisensory modes could enhance performance. One of the suggested methods for this is multisensory combination, which enhances multisensory perception with the complementary features of sensory information. The present thesis argued that multisensory combination could be applied to sensory substitution techniques as it was evidenced that tactile and auditory sensory substitution carry axis specific information resolution. In other words, the axis specific sensory information can be complimented via multisensory combination for robust perceptual judgements. By doing so, not only could some of the handicaps behind the widespread adoption of sensory substitution techniques can be addressed, but also mainstream cross-modal displays that appeal to a wider range of people could be developed.

In the scientific literature, there is a vast amount of sensory substitution techniques with distinct methods of transforming sensory signals. The current thesis utilised two of the commercially available cross-modal displays and evaluated their performance with respect to multisensory combination. This availability brings an advantage to the users and researchers so that they can try and test out the use cases described here in a wider context. The findings, however, may not yet be generalised to all sensory substitution techniques. This further requires scientific confirmation across different bodies of research that investigate the information capacity of sensory substitution techniques and their perceived resolutions. The methodologies described in the present thesis might offer guidance in this quest. Other researchers have also recently benchmarked the performance of different sensory substitution techniques to enhance their display modes [17]. Considering the state-of-the-art knowledge of sensory substitution and their information capacity/resolution, future research should

therefore examine which sensory and also cognitive representations form the basis of multisensory perception with cross-modal displays. This would not only contribute significantly to our understanding of cross-modal cognition but also tremendously enhance how we build and interact with technology. Indeed, in the recent years, the concepts of cross-modal cognition have spread across different interdisciplinary research, such as neural networks, artificial intelligence and cognitive robotics (e.g. [4,10,15,16]).

One of the most fundamental research questions sensory substitution raises is whether the perception of sensory information has a singular cognitive, perhaps neural, elementary basis. This origin could then be moulded and represented distinctively via various qualia. Sensory substitution then implies more than just the phenomenological substitution between the qualia of experiencing sensory information. Recent theories of the neocortex in neuroscience, in this regard, attempt to unify how our sensory information processing might share the same neural circuitry (e.g. [8,9,11]). If these theories hold true, sensory substitution techniques could be applied as 'a universal brain-computer interface' [5]. That is, being able to make sense of information, which is inaccessible to our natural sensory organs, would be possible using the way the brain can inclusively process information regardless of its format. This would lead to unprecedented applications and implications of inclusive design in how we interact with technology and how technology interacts with us in the future. Perhaps, in this context, sensory substitution can be considered to be the cognitive transmutation of information.

References

1. Gabriel Arnold, Jacques Pesnot-Lerousseau, and Malika Auvray. 2017. Individual Differences in Sensory Substitution. *Multisensory Research* 30, 6: 579–600. https://doi.org/10.1163/22134808-00002561
2. Malika Auvray and Mirko Farina. 2017. Patrolling the Boundaries of Synaesthesia. In *Synaesthesia: Philosophical & Psychological Challenges*, O Deroy (ed.). Oxford University Press, Oxford, 248–274.
3. Malika Auvray and Erik Myin. 2009. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science* 33, 6: 1036–1058. https://doi.org/10.1111/j.1551-6709.2009.01040.x
4. Tadeo Corradi, Peter Hall, and Pejman Iravani. 2017. Object recognition combining vision and touch. *Robotics and Biomimetics* 4, 1: 2. https://doi.org/10.1186/s40638-017-0058-2
5. Yuri Danilov and Mitchell Tyler. 2005. BrainPort: An Alternative Input to the Brain. *Journal of Integrative Neuroscience* 04, 04: 537–550. https://doi.org/10.1142/S0219635205000914
6. Ophelia Deroy and Malika Auvray. 2012. Reading the World through the Skin and Ears: A New Perspective on Sensory Substitution. *Frontiers in Psychology* 3. https://doi.org/10.3389/fpsyg.2012.00457
7. Ophelia Deroy and Malika Auvray. 2014. A Crossmodal Perspective on Sensory Substitution. In *Perception and Its Modalities*. Oxford University Press, 327–349. https://doi.org/10.1093/acprof:oso/9780199832798.003.0014
8. Jeff Hawkins and Subutai Ahmad. 2016. Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex. *Frontiers in Neural Circuits* 10. https://doi.org/10.3389/fncir.2016.00023
9. Jeff Hawkins, Subutai Ahmad, and Yuwei Cui. 2017. A Theory of How Columns in the Neocortex Enable Learning the Structure of the World. *Frontiers in Neural Circuits* 11. https://doi.org/10.3389/fncir.2017.00081
10. Jeff Hawkins and Sandra. Blakeslee. 2005. *On intelligence*. Henry Holt and Co, New York.
11. Jeff Hawkins, Marcus Lewis, Mirko Klukas, Scott Purdy, and Subutai Ahmad. 2019. A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex. *Frontiers in Neural Circuits* 12. https://doi.org/10.3389/fncir.2018.00121
12. C. Lenay, S. Canu, and P. Villon. 1997. Technology and perception: the contribution of sensory substitution systems. In *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age*, 44–53. https://doi.org/10.1109/CT.1997.617681
13. Charles Lenay and Gunnar Declerck. 2018. Technologies to Access Space Without Vision. Some Empirical Facts and Guiding Theoretical Principles. In *Mobility of Visually Impaired People*. Springer International Publishing, Cham, 53–75. https://doi.org/10.1007/978-3-319-54446-5_2
14. Charles Lenay, Olivier Gapenne, Sylvain Hanneton, Catherine Marque, and Christelle Genouëlle. 2003. SENSORY SUBSTITUTION: LIMITS AND PERSPECTIVES. In *Touching for Knowing: Cognitive Psychology of Haptic Manual Perception*, Y Hatwell, A Streri and E Gentaz (eds.). John Benjamins Publishing Company , Amsterdam, Netherlands, 275–292.
15. Yunzhu Li, Jun-Yan Zhu, Russ Tedrake, and Antonio Torralba. 2019. Connecting Touch and Vision via Cross-Modal Prediction. In *CVPR*.
16. Alessandro Di Nuovo and Angelo Cangelosi. 2015. Artificial Mental Imagery in Cognitive Robots Interaction. In *2015 IEEE Symposium Series on Computational Intelligence*, 91–96. https://doi.org/10.1109/SSCI.2015.23
17. Michael Richardson, Jan Thar, James Alvarez, Jan Borchers, Jamie Ward, and Giles Hamilton-Fletcher. 2019. How Much Spatial Information Is Lost in the Sensory Substitution Process? Comparing Visual, Tactile, and Auditory Approaches. *Perception* 48, 11: 1079–1103. https://doi.org/10.1177/0301006619873194